

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2004 年 12 月 16 日 (16.12.2004)

PCT

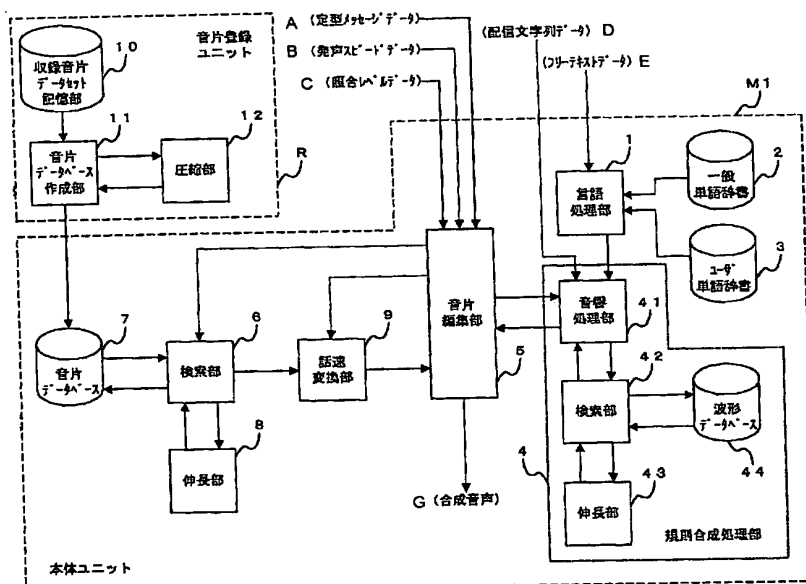
(10) 国際公開番号
WO 2004/109659 A1

- (51) 国際特許分類⁷: G10L 13/06 [JP/JP]; 〒1928525 東京都八王子市石川町2967-3 Tokyo (JP).
- (21) 国際出願番号: PCT/JP2004/008087
- (22) 国際出願日: 2004 年 6 月 3 日 (03.06.2004)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願2003-160657 2003 年 6 月 5 日 (05.06.2003) JP
特願2004-142907 2004 年 4 月 9 日 (09.04.2004) JP
特願2004-142906 2004 年 4 月 9 日 (09.04.2004) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社 ケンウッド (KABUSHIKI KAISHA KENWOOD)
- (72) 発明者; および
(75) 発明者/出願人 (米国についてののみ): 佐藤 寧 (SATO, Yasushi) [JP/JP]; 〒2700163 千葉県流山市南流山 4-16-18 ローヤルコーポ・ワタナベ 401 Chiba (JP).
- (74) 代理人: 岡部 正夫, 外 (OKABE, Masao et al.); 〒1000005 東京都千代田区丸の内 3-2-3 富士ビル 602号室 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM,

[続葉有]

(54) Title: SPEECH SYNTHESIS DEVICE, SPEECH SYNTHESIS METHOD, AND PROGRAM

(54) 発明の名称: 音声合成装置、音声合成方法及びプログラム



10...COLLECTED VOICE PIECE DATASET STORAGE SECTION
R...VOICE PIECE REGISTRATION UNIT
11...VOICE PIECE DATABASE CREATION SECTION
12...COMPRESSION SECTION
A...(MESSAGE TEMPLATE DATA)
B...(SPEECH SPEED DATA)
C...(COLLATING LEVEL DATA)
D...(DELIVERY CHARACTER STRING DATA)
E...(FREE TEXT DATA)
M1...BODY UNIT
7...VOICE PIECE DATABASE
6...SEARCH SECTION

8...DECOMPRESSION SECTION
9...SPEECH SPEED CONVERTER
5...VOICE PIECE EDITOR
G...(SYNTHETIC SPEECH)
1...LANGUAGE PROCESSOR
2...GENERAL WORD DICTIONARY
3...USER WORD DICTIONARY
41...ACOUSTIC PROCESSOR
42...SEARCH SECTION
43...WAVEFORM DATABASE
44...DECOMPRESSION SECTION
4...RULE COMBINATION PROCESSOR

(57) Abstract: A simply configured speech synthesis device and the like for producing a natural synthetic speech at high speed. When data representing a message template is supplied, a voice piece editor (5) searches a voice piece database (7) for voice piece data on a voice piece whose sound matches a voice piece in the message template. Further, the voice piece editor (5) predicts the cadence of the message template and selects, one at a time, a best match of each voice piece in the message template from the voice piece data that has been retrieved, according to the cadence prediction result. For a voice piece for which no match can be selected, an acoustic processor (41) is instructed to supply waveform data representing the waveform of each unit voice. The voice piece data that is selected and the waveform data that is supplied by the acoustic processor (41) are combined to generate data representing a synthetic speech.

(57) 要約: 簡単な構成で高速に自然な合成音声を得るための音声合成装置等を提供することで

ある。定型メッセージを表すデータが供給されると、音声編集部5は、定型メッセージ

[続葉有]



DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告書

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG,

2 文字コード及び他の略語については、定期発行される各 *PCT* ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

内の音片と読みが合致する音片の音片データを音片データベース 7 から索出させる。一方で音片編集部 5 は定型メッセージの韻律予測を行い、索出された音片データのうちから定型メッセージ内の各音片に最もよく合致するものを 1 個ずつ、韻律予測結果に基づいて選択する。選択できなかった音片については音響処理部 4 1 に単位音声毎の波形を表す波形データを供給させる。そして、選択した音片データや、音響処理部 4 1 に供給させた波形データを互いに結合して、合成音声を表すデータを生成する。

明 細 書

音声合成装置、音声合成方法及びプログラム

技術分野

この発明は、音声合成装置、音声合成方法及びプログラムに関する。

背景技術

音声合成する手法として、録音編集方式と呼ばれる手法がある。録音編集方式は、駅の音声案内システムや、車載用のナビゲーション装置などに用いられている。

録音編集方式は、単語と、この単語を読み上げる音声を表す音声データとを対応付けておき、音声合成する対象の文章を単語に区切ってから、これらの単語に対応付けられた音声データを取得してつなぎ合わせる、という手法である（例えば、特開平 1 0 - 4 9 1 9 3 号公報参照）。

発明の開示

しかし、音声データを単につなぎ合わせた場合、音声データ同士の境界では通常、音声のピッチ成分の周波数が不連続的に変化する、等の理由で、合成音声が不自然なものとなる。

この問題を解決する手法としては、同一の音素を互いに異なった韻律で読み上げる音声を表す複数の音声データを用意し、一方で音声合成する対象の文章に韻律予測を施して、予測結果に合致する音声データを選び出してつなぎ合わせる、という手法が考えられる。

しかし、音声データを音素毎に用意して録音編集方式により自然な合成音声を得ようとする、と、音声データを記憶する記憶装置には

膨大な記憶容量が必要となる。また、検索する対象のデータの量も膨大なものとなる。

この発明は、上記実状に鑑みてなされたものであり、簡単な構成で高速に自然な合成音声を得るための音声合成装置、音声合成方法及びプログラムを提供することを目的とする。

上記目的を達成するため、この発明の第 1 の観点にかかる音声合成装置は、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

より構成されることを特徴とする。

また、この発明の第 2 の観点にかかる音声合成装置は、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音片の波形を表す音声データを合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

より構成されることを特徴とする。

前記選択手段は、韻律が韻律予測結果に前記所定の条件下で合致しない音片データを、選択の対象から除外するものであってもよい。

前記欠落部分合成手段は、

音素を表し、又は、音素を構成する素片を表すデータを複数記憶する記憶手段と、

前記選択手段が音片データを選択できなかった前記音声に含まれる音素を特定し、特定した音素又は当該音素を構成する素片を表すデータを前記記憶手段より取得して互いに結合することにより、当該音声の波形を表す音声データを合成する合成手段と、を備えるものであってもよい。

前記欠落部分合成手段は、前記選択手段が音片データを選択できなかった前記音声の韻律を予測する欠落部分韻律予測手段を備えてもよく、

前記合成手段は、前記選択手段が音片データを選択できなかった前記音声に含まれる音素を特定し、特定した音素又は当該音素を構成する素片を表すデータを前記記憶手段より取得し、取得したデータを、当該データが表す音素又は素片が、前記欠落部分韻律予測手段による韻律の予測結果に合致するように変換して、変換されたデ

ータを互いに結合することにより、当該音声の波形を表す音声データを合成するものであってもよい。

前記欠落部分合成手段は、前記韻律予測手段が予測した韻律に基づいて、前記選択手段が音片データを選択できなかった音声について、当該音片の波形を表す音声データを合成するものであってもよい。

前記音片記憶手段は、音片データが表す音片のピッチの時間変化を表す韻律データを、当該音片データに対応付けて記憶していてもよく、

前記選択手段は、各前記音片データのうちから、前記文章を構成する音声と読みが共通しており、且つ、対応付けられている韻律データが表すピッチの時間変化が韻律の予測結果に最も近い音片データを選択するものであってもよい。

前記音声合成装置は、前記合成音声を発声するスピードの条件を指定する発声スピードデータを取得し、前記合成音声を表すデータを構成する音片データ及び／又は音声データを、当該発声スピードデータが指定する条件を満たすスピードで発話される音声を表すように選択又は変換する発話スピード変換手段を備えるものであってもよい。

前記発話スピード変換手段は、前記合成音声を表すデータを構成する音片データ及び／又は音声データから素片を表す区間を除去し、又は、当該音片データ及び／又は音声データに素片を表す区間を追加することによって、当該音片データ及び／又は音声データを、前記発声スピードデータが指定する条件を満たすスピードで発話される音声を表すよう変換するものであってもよい。

前記音片記憶手段は、音片データの読みを表す表音データを、当該音片データに対応付けて記憶していてもよく、

前記選択手段は、前記文章を構成する音声の読みに合致する読みを表す表音データが対応付けられている音片データを、当該音声と読みが共通する音片データとして扱うものであってもよい。

また、この発明の第 3 の観点にかかる音声合成方法は、

音片を表す音片データを複数記憶し、

文章を表す文章情報を入力し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択し、

前記文章を構成する音声のうち、音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成し、

選択した音片データ及び合成した音声データを互いに結合することにより、合成音声を表すデータを生成する、

ことを特徴とする。

また、この発明の第 4 の観点にかかる音声合成方法は、

音片を表す音片データを複数記憶し、

文章を表す文章情報を入力して、当該文章を構成する音声の韻律を予測し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択し、

前記文章を構成する音声のうち、音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成し、

選択した音片データ及び合成した音声データを互いに結合するこ

- 6 -

とにより、合成音声を表すデータを生成する、

ことを特徴とする。

また、この発明の第 5 の観点にかかるプログラムは、

コンピュータを、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

して機能させるためのものであることを特徴とする。

また、この発明の第 6 の観点にかかるプログラムは、

コンピュータを、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音声の波形を表す音声データを

合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

して機能させるためのものであることを特徴とする。

上記目的を達成するため、この発明の第 7 の観点にかかる音声合成装置は、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に最も近い音片データを選択する選択手段と、

選択された音片データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

より構成されることを特徴とする。

前記選択手段は、韻律が韻律予測結果に所定の条件下で合致しない音片データを、選択の対象から除外するものであってもよい。

前記音声合成装置は、前記合成音声を発声するスピードの条件を指定する発声スピードデータを取得し、前記合成音声を表すデータを構成する音片データ及び／又は音声データを、当該発声スピードデータが指定する条件を満たすスピードで発話される音声を表すように選択又は変換する発話スピード変換手段を備えるものであってもよい。

前記発話スピード変換手段は、前記合成音声を表すデータを構成

する音片データ及び／又は音声データから素片を表す区間を除去し、又は、当該音片データ及び／又は音声データに素片を表す区間を追加することによって、当該音片データ及び／又は音声データを、前記発声スピードデータが指定する条件を満たすスピードで発話される音声を表すよう変換するものであってもよい。

前記音片記憶手段は、音片データが表す音片のピッチの時間変化を表す韻律データを、当該音片データに対応付けて記憶していてもよく、

前記選択手段は、各前記音片データのうちから、前記文章を構成する音声と読みが共通しており、且つ、対応付けられている韻律データが表すピッチの時間変化が韻律の予測結果に最も近い音片データを選択するものであってもよい。

前記音片記憶手段は、音片データの読みを表す表音データを、当該音片データに対応付けて記憶していてもよく、

前記選択手段は、前記文章を構成する音声の読みに合致する読みを表す表音データが対応付けられている音片データを、当該音声と読みが共通する音片データとして扱うものであってもよい。

また、この発明の第 8 の観点にかかる音声合成方法は、

音片を表す音片データを複数記憶し、

文章を表す文章情報を入力して、当該文章を構成する音声の韻律を予測し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に最も近い音片データを選択し、

選択された音片データを互いに結合することにより、合成音声を

表すデータを生成する、

ことを特徴とする。

また、この発明の第 9 の観点にかかるプログラムは、

コンピュータを、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に最も近い音片データを選択する選択手段と、

選択された音片データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

して機能させるためのものであることを特徴とする。

以上説明したように、この発明によれば、簡単な構成で高速に自然な合成音声を得るための音声合成装置、音声合成方法及びプログラムが実現される。

図面の簡単な説明

第 1 図は、この発明の第 1 の実施の形態に係る音声合成システムの構成を示すブロック図である。

第 2 図は、音片データベースのデータ構造を模式的に示す図である。

第 3 図は、この発明の第 2 の実施の形態に係る音声合成システムの構成を示すブロック図である。

第 4 図は、この発明の第 1 の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータがフリーテキストデータを取

得した場合の処理を示すフローチャートである。

第 5 図は、この発明の第 1 の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

第 6 図は、この発明の第 1 の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

第 7 図は、第 3 図の本体ユニットの機能を行うパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

第 8 図は、第 3 図の本体ユニットの機能を行うパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

第 9 図は、第 3 図の本体ユニットの機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

発明を実施するための最良の形態

以下、図面を参照して、この発明の実施の形態を説明する。

(第 1 の実施の形態)

第 1 図は、この発明の第 1 の実施の形態に係る音声合成システムの構成を示す図である。図示するように、この音声合成システムは、本体ユニット M 1 と、音片登録ユニット R とにより構成されている。

本体ユニット M 1 は、言語処理部 1 と、一般単語辞書 2 と、ユーザ単語辞書 3 と、規則合成処理部 4 と、音片編集部 5 と、検索部 6

と、音片データベース 7 と、伸長部 8 と、話速変換部 9 とにより構成されている。このうち、規則合成処理部 4 は、音響処理部 4 1 と、検索部 4 2 と、伸長部 4 3 と、波形データベース 4 4 とより構成されている。

言語処理部 1、音響処理部 4 1、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 は、いずれも、C P U (Central Processing Unit) や D S P (Digital Signal Processor) 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。

なお、言語処理部 1、音響処理部 4 1、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。従って、例えば、伸長部 4 3 の機能を行うプロセッサが伸長部 8 の機能を行ってもよいし、1 個のプロセッサが音響処理部 4 1、検索部 4 2 及び伸長部 4 3 の機能を兼ねて行ってもよい。

一般単語辞書 2 は、P R O M (Programmable Read Only Memory) やハードディスク装置等の不揮発性メモリより構成されている。一般単語辞書 2 には、表意文字 (例えば、漢字など) を含む単語等と、この単語等の読みを表す表音文字 (例えば、カナや発音記号など) とが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。

ユーザ単語辞書 3 は、E E P R O M (Electrically Erasable/Programmable Read Only Memory) やハードディスク装置等のデータ書き換え可能な不揮発性メモリと、この不揮発性メモ

リへのデータの書き込みを制御する制御回路とにより構成されている。なお、プロセッサがこの制御回路の機能を行ってもよく、言語処理部 1、音響処理部 4 1、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 の一部又は全部の機能を行うプロセッサがユーザ単語辞書 3 の制御回路の機能を行うようにしてもよい。

ユーザ単語辞書 3 は、表意文字を含む単語等と、この単語等の読みを表す表音文字とを、ユーザの操作に従って外部より取得し、互いに対応付けて記憶する。ユーザ単語辞書 3 には、一般単語辞書 2 に記憶されていない単語等とその読みを表す表音文字とが格納されていれば十分である。

波形データベース 4 4 は、PROM やハードディスク装置等の不揮発性メモリより構成されている。波形データベース 4 4 には、表音文字と、この表音文字が表す単位音声の波形を表す波形データをエントロピー符号化して得られる圧縮波形データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。単位音声は、規則合成方式の手法で用いられる程度の短い音声であり、具体的には、音素や、V C V (Vowel-Consonant-Vowel) 音節などの単位で区切られる音声である。なお、エントロピー符号化される前の波形データは、例えば、PCM (Pulse Code Modulation) されたデジタル形式のデータからなっていればよい。

音片データベース 7 は、PROM やハードディスク装置等の不揮発性メモリより構成されている。

音片データベース 7 には、例えば、第 2 図に示すデータ構造を有

するデータが記憶されている。すなわち、図示するように、音片データベース 7 に格納されているデータは、ヘッダ部 H D R、インデックス部 I D X、ディレクトリ部 D I R 及びデータ部 D A T の 4 種に分かれている。

なお、音片データベース 7 へのデータの格納は、例えば、この音声合成システムの製造者によりあらかじめ行われ、及び／又は、音片登録ユニット R が後述する動作を行うことにより行われる。

ヘッダ部 H D R には、音片データベース 7 を識別するデータや、インデックス部 I D X、ディレクトリ部 D I R 及びデータ部 D A T のデータ量、データの形式、著作権等の帰属などを示すデータが格納される。

データ部 D A T には、音片の波形を表す音片データをエントロピー符号化して得られる圧縮音片データが格納されている。

なお、音片とは、音声のうち音素 1 個以上を含む連続した 1 区間をいい、通常は単語 1 個分又は複数個分の区間からなる。音片は接続詞を含む場合もある。

また、エントロピー符号化される前の音片データは、上述の圧縮波形データの生成のためエントロピー符号化される前の波形データと同じ形式のデータ（例えば、P C M されたデジタル形式のデータ）からなっていればよい。

ディレクトリ部 D I R には、個々の圧縮音声データについて、

(A) この圧縮音片データが表す音片の読みを示す表音文字を表すデータ（音片読みデータ）、

(B) この圧縮音片データが格納されている記憶位置の先頭のアドレスを表すデータ、

- (C) この圧縮音片データのデータ長を表すデータ、
- (D) この圧縮音片データが表す音片の発声スピード（再生した場合の時間長）を表すデータ（スピード初期値データ）、
- (E) この音片のピッチ成分の周波数の時間変化を表すデータ（ピッチ成分データ）、

が、互いに対応付けられた形で格納されている。（なお、音片データベース 7 の記憶領域にはアドレスが付されているものとする。）

なお、第 2 図は、データ部 D A T に含まれるデータとして、読みが「サイタマ」である音片の波形を表す、データ量 1 4 1 0 h バイトの圧縮音片データが、アドレス 0 0 1 A 3 6 A 6 h を先頭とする論理的位置に格納されている場合を例示している。（なお、本明細書及び図面において、末尾に“h”を付した数字は 1 6 進数を表す。）

なお、上述の (A) ~ (E) のデータの集合のうち少なくとも (A) のデータ（すなわち音片読みデータ）は、音片読みデータが表す表音文字に基づいて決められた順位に従ってソートされた状態で（例えば、表音文字がカナであれば、五十音順に従って、アドレス降順に並んだ状態で）、音片データベース 7 の記憶領域に格納されている。

また、上述のピッチ成分データは、例えば、図示するように、音片のピッチ成分の周波数を音片の先頭からの経過時間の 1 次関数で近似した場合における、この 1 次関数の切片 β 及び勾配 α の値を示すデータからなっていればよい。（勾配 α の単位は例えば [ヘルツ / 秒] であればよく、切片 β の単位は例えば [ヘルツ] であればよい。）

また、ピッチ成分データには更に、圧縮音片データが表す音片が鼻濁音化されているか否か、及び、無声化されているか否かを表す

図示しないデータも含まれているものとする。

インデックス部 I D X には、ディレクトリ部 D I R のデータのおおよその論理的位置を音片読みデータに基づいて特定するためのデータが格納されている。具体的には、例えば、音片読みデータがカナを表すものであるとして、カナ文字と、先頭 1 字がこのカナ文字であるような音片読みデータがどのような範囲のアドレスにあるかを示すデータ（ディレクトリアドレス）とが、互いに対応付けて格納されている。

なお、一般単語辞書 2、ユーザ単語辞書 3、波形データベース 4 及び音片データベース 7 の一部又は全部の機能を単一の不揮発性メモリが行うようにしてもよい。

音片登録ユニット R は、図示するように、収録音片データセット記憶部 10 と、音片データベース作成部 11 と、圧縮部 12 とにより構成されている。なお、音片登録ユニット R は音片データベース 7 とは着脱可能に接続されていてもよく、この場合は、音片データベース 7 に新たにデータを書き込むときを除いては、音片登録ユニット R を本体ユニット M 1 から切り離した状態で本体ユニット M 1 に後述の動作を行わせてよい。

収録音片データセット記憶部 10 は、ハードディスク装置等のデータ書き換え可能な不揮発性メモリより構成されている。

収録音片データセット記憶部 10 には、音片の読みを表す表音文字と、この音片を人が実際に発声したものを集音して得た波形を表す音片データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。なお、この音片データは、例えば、PCM されたデジタル形式のデータからなっていれば

よい。

音片データベース作成部 11 及び圧縮部 12 は、CPU等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、このプログラムに従って後述する処理を行う。

なお、音片データベース作成部 11 及び圧縮部 12 の一部又は全部の機能を単一のプロセッサが行うようにしてもよく、また、言語処理部 1、音響処理部 41、検索部 42、伸長部 43、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 の一部又は全部の機能を行うプロセッサが音片データベース作成部 11 や圧縮部 12 の機能を更に行ってもよい。また、音片データベース作成部 11 や圧縮部 12 の機能を行うプロセッサが、収録音片データセット記憶部 10 の制御回路の機能を兼ねてもよい。

音片データベース作成部 11 は、収録音片データセット記憶部 10 より、互いに対応付けられている表音文字及び音片データを読み出し、この音片データが表す音声のピッチ成分の周波数の時間変化と、発声スピードとを特定する。

発声スピードの特定は、例えば、この音片データのサンプル数を数えることにより特定すればよい。

一方、ピッチ成分の周波数の時間変化は、例えば、この音片データにケプストラム解析を施すことにより特定すればよい。具体的には、例えば、音片データが表す波形を時間軸上で多数の小部分へと区切り、得られたそれぞれの小部分の強度を、元の値の対数（対数の底は任意）に実質的に等しい値へと変換し、値が変換されたこの小部分のスペクトル（すなわち、ケプストラム）を、高速フーリエ

変換の手法（あるいは、離散変数をフーリエ変換した結果を表すデータを生成する他の任意の手法）により求める。そして、このケプストラムの極大値を与える周波数のうちの最小値を、この小部分におけるピッチ成分の周波数として特定する。

なお、ピッチ成分の周波数の時間変化は、例えば、特開 2003-108172 号公報に開示された手法に従って音片データをピッチ波形データへと変換してから、このピッチ波形データに基づいて特定するようにすると良好な結果が期待できる。具体的には、音片データをフィルタリングしてピッチ信号を抽出し、抽出されたピッチ信号に基づいて、音片データが表す波形を単位ピッチ長の区間へと区切り、各区間について、ピッチ信号との相関関係に基づいて位相のずれを特定して各区間の位相を揃えることにより、音片データをピッチ波形信号へと変換すればよい。そして、得られたピッチ波形信号を音片データとして扱い、ケプストラム解析を行う等することにより、ピッチ成分の周波数の時間変化を特定すればよい。

一方、音片データベース作成部 11 は、収録音片データセット記憶部 10 より読み出した音片データを圧縮部 12 に供給する。

圧縮部 12 は、音片データベース作成部 11 より供給された音片データをエントロピー符号化して圧縮音片データを作成し、音片データベース作成部 11 に返送する。

音片データの発声スピード及びピッチ成分の周波数の時間変化を特定し、この音片データがエントロピー符号化され圧縮音片データとなって圧縮部 12 より返送されると、音片データベース作成部 11 は、この圧縮音片データを、データ部 DAT を構成するデータとして、音片データベース 7 の記憶領域に書き込む。

また、音片データベース作成部 11 は、書き込んだ圧縮音片データが表す音片の読みを示すものとして収録音片データセット記憶部 10 より読み出した表音文字を、音片読みデータとして音片データベース 7 の記憶領域に書き込む。

また、書き込んだ圧縮音片データの、音片データベース 7 の記憶領域内での先頭のアドレスを特定し、このアドレスを上述の (B) のデータとして音片データベース 7 の記憶領域に書き込む。

また、この圧縮音片データのデータ長を特定し、特定したデータ長を、(C) のデータとして音片データベース 7 の記憶領域に書き込む。

また、この圧縮音片データが表す音片の発声スピード及びピッチ成分の周波数の時間変化を特定した結果を示すデータを生成し、スピード初期値データ及びピッチ成分データとして音片データベース 7 の記憶領域に書き込む。

次に、この音声合成システムの動作を説明する。

まず、言語処理部 1 が、この音声合成システムに音声を合成させる対象としてユーザが用意した、表意文字を含む文章（フリーテキスト）を記述したフリーテキストデータを外部から取得したとして説明する。

なお、言語処理部 1 がフリーテキストデータを取得する手法は任意であり、例えば、図示しないインターフェース回路を介して外部の装置やネットワークから取得してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体（例えば、フロッピー（登録商標）ディスクや CD-ROM など）から、この記録媒体ドライブ装置を介して読み取ってもよい。

また、言語処理部 1 の機能を行っているプロセッサが、自ら実行している他の処理で用いたテキストデータを、フリーテキストデータとして、言語処理部 1 の処理へと引き渡すようにしてもよい。

プロセッサが実行する当該他の処理としては、例えば、音声を表す音声データを取得し、この音声データに音声認識を施すことにより、この音声を表す語句を特定し、特定した語句に基づいて、この音声の発話者の要求の内容を特定して、特定した要求を満足させるために実行すべき処理を特定して実行するようなエージェント装置の機能をプロセッサに行わせるための処理などが考えられる。

フリーテキストデータを取得すると、言語処理部 1 は、このフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書 2 やユーザ単語辞書 3 を検索することにより特定する。そして、この表意文字を、特定した表音文字へと置換する。そして、言語処理部 1 は、フリーテキスト内の表意文字がすべて表音文字へと置換した結果得られる表音文字列を、音響処理部 4 1 へと供給する。

音響処理部 4 1 は、言語処理部 1 より表音文字列を供給されると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字を表す単位音声の波形を検索するよう、検索部 4 2 に指示する。

検索部 4 2 は、この指示に応答して波形データベース 4 4 を検索し、表音文字列に含まれるそれぞれの表音文字を表す単位音声の波形を表す圧縮波形データを索出する。そして、索出された圧縮波形データを伸長部 4 3 へと供給する。

伸長部 4 3 は、検索部 4 2 より供給された圧縮波形データを、圧縮される前の波形データへと復元し、検索部 4 2 へと返送する。検

索部 4 2 は、伸長部 4 3 より返送された波形データを、検索結果として音響処理部 4 1 へと供給する。

音響処理部 4 1 は、検索部 4 2 より供給された波形データを、言語処理部 1 より供給された表音文字列内での各表音文字の並びに従った順序で、音片編集部 5 へと供給する。

音片編集部 5 は、音響処理部 4 1 より波形データを供給されると、この波形データを、供給された順序で互いに結合し、合成音声を表すデータ（合成音声データ）として出力する。フリーテキストデータに基づいて合成されたこの合成音声は、規則合成方式の手法により合成された音声に相当する。

なお、音片編集部 5 が合成音声データを出力する手法は任意であり、例えば、図示しない D/A（Digital-to-Analog）変換器やスピーカを介して、この合成音声データが表す合成音声を再生するようにしてもよい。また、図示しないインターフェース回路を介して外部の装置やネットワークに送出してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体へ、この記録媒体ドライブ装置を介して書き込んでもよい。また、音片編集部 5 の機能を行っているプロセッサが、自ら実行している他の処理へと、合成音声データを引き渡すようにしてもよい。

次に、音響処理部 4 1 が、外部より配信された、表音文字列を表すデータ（配信文字列データ）を取得したとする。（なお、音響処理部 4 1 が配信文字列データを取得する手法も任意であり、例えば、言語処理部 1 がフリーテキストデータを取得する手法と同様の手法で配信文字列データを取得すればよい。）

この場合、音響処理部 4 1 は、配信文字列データが表す表音文字

列を、言語処理部 1 より供給された表音文字列と同様に扱う。この結果、配信文字列データが表す表音文字列に含まれる表音文字に対応する圧縮波形データが検索部 4 2 により索出され、圧縮される前の波形データが伸長部 4 3 により復元される。復元された各波形データは音響処理部 4 1 を介して音片編集部 5 へと供給され、音片編集部 5 が、この波形データを、配信文字列データが表す表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する。配信文字列データに基づいて合成されたこの合成音声データも、規則合成方式の手法により合成された音声を表す。

次に、音片編集部 5 が、定型メッセージデータ、発声スピードデータ、及び照合レベルデータを取得したとする。

なお、定型メッセージデータは、定型メッセージを表音文字列として表すデータであり、発声スピードデータは、定型メッセージデータが表す定型メッセージの発声スピードの指定値（この定型メッセージを発声する時間長の指定値）を示すデータである。照合レベルデータは、検索部 6 が行う後述の検索処理における検索条件を指定するデータであり、以下では「1」、「2」又は「3」のいずれかの値をとるものとし、「3」が最も厳格な検索条件を示すものとする。

また、音片編集部 5 が定型メッセージデータや発声スピードデータや照合レベルデータを取得する手法は任意であり、例えば、言語処理部 1 がフリーテキストデータを取得する手法と同様の手法で定型メッセージデータや発声スピードデータや照合レベルデータを取得すればよい。

定型メッセージデータ、発声スピードデータ、及び照合レベルデ

ータが音片編集部 5 に供給されると、音片編集部 5 は、定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出するよう、検索部 6 に指示する。

検索部 6 は、音片編集部 5 の指示に応答して音片データベース 7 を検索し、該当する圧縮音片データと、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データとを索出し、索出された圧縮波形データを伸長部 4 3 へと供給する。複数の圧縮音片データが共通の表音文字ないし表音文字列に該当する場合も、該当する圧縮音片データすべてが、音声合成に用いられるデータの候補として索出される。一方、圧縮音片データを索出できなかった音片があった場合、検索部 6 は、該当する音片を識別するデータ（以下、欠落部分識別データと呼ぶ）を生成する。

伸長部 4 3 は、検索部 6 より供給された圧縮音片データを、圧縮される前の音片データへと復元し、検索部 6 へと返送する。検索部 6 は、伸長部 4 3 より返送された音片データと、索出された音片読みデータ、スピード初期値データ及びピッチ成分データとを、検索結果として話速変換部 9 へと供給する。また、欠落部分識別データを生成した場合は、この欠落部分識別データも話速変換部 9 へと供給する。

一方、音片編集部 5 は、話速変換部 9 に対し、話速変換部 9 に供給された音片データを変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致するようにすることを指示する。

話速変換部 9 は、音片編集部 5 の指示に応答し、検索部 6 より供給された音片データを指示に合致するように変換して、音片編集部 5 に供給する。具体的には、例えば、検索部 6 より供給された音片データの元の時間長を、索出されたスピード初期値データに基づいて特定した上、この音片データをリサンプリングして、この音片データのサンプル数を、音片編集部 5 の指示したスピードに合致する時間長にすればよい。

また、話速変換部 9 は、検索部 6 より供給された音片読みデータ及びピッチ成分データも音片編集部 5 に供給し、欠落部分識別データを検索部 6 より供給された場合は、更にこの欠落部分識別データも音片編集部 5 に供給する。

なお、発声スピードデータが音片編集部 5 に供給されていない場合、音片編集部 5 は、話速変換部 9 に対し、話速変換部 9 に供給された音片データを変換せずに音片編集部 5 に供給するよう指示すればよく、話速変換部 9 は、この指示に応答し、検索部 6 より供給された音片データをそのまま音片編集部 5 に供給すればよい。

音片編集部 5 は、話速変換部 9 より音片データ、音片読みデータ及びピッチ成分データを供給されると、供給された音片データのうちから、定型メッセージを構成する音片の波形に近似できる波形を表す音片データを、音片 1 個につき 1 個ずつ選択する。ただし、音片編集部 5 は、いかなる条件を満たす波形を定型メッセージの音片に近い波形とするかを、取得した照合レベルデータに従って設定する。

具体的には、まず、音片編集部 5 は、定型メッセージデータが表す定型メッセージに、例えば「藤崎モデル」や「T o B I (Tone and

Break Indices)」等の韻律予測の手法に基づいた解析を加えることにより、この定型メッセージの韻律（アクセント、イントネーション、強勢、音素の時間長など）を予測する。

次に、音片編集部 5 は、例えば、

(1) 照合レベルデータの値が「1」である場合は、話速変換部 9 より供給された音片データ（すなわち、定型メッセージ内の音片と読みが合致する音片データ）をすべて、定型メッセージ内の音片の波形に近いものとして選択する。

(2) 照合レベルデータの値が「2」である場合は、(1) の条件（つまり、読みを表す表音文字の合致という条件）を満たし、更に、音片データのピッチ成分の周波数の時間変化を表すピッチ成分データの内容と定型メッセージに含まれる音片のアクセント（いわゆる韻律）の予測結果との間に所定量以上の強い相関がある場合（例えば、アクセントの位置の時間差が所定量以下である場合）に限り、この音片データが定型メッセージ内の音片の波形に近いものとして選択する。なお、定型メッセージ内の音片のアクセントの予測結果は、定型メッセージの韻律の予測結果より特定できるものであり、音片編集部 5 は、例えば、ピッチ成分の周波数が最も高いと予測されている位置をアクセントの予測位置であると解釈すればよい。一方、音片データが表す音片のアクセントの位置については、例えば、ピッチ成分の周波数が最も高い位置を上述のピッチ成分データに基づいて特定し、この位置をアクセントの位置であると解釈すればよい。また、韻律予測は、文章全体に対して行ってもよいし、文章を所定の単位に分割し、それぞれの単位に対して行ってもよい。

(3) 照合レベルデータの値が「3」である場合は、(2) の条件

(つまり、読みを表す表音文字及びアクセントの合致という条件)を満たし、更に、音片データが表す音声の鼻濁音化や無声化の有無が、定型メッセージの韻律の予測結果に合致している場合に限り、この音片データが定型メッセージ内の音片の波形に近いものとして選択する。音片編集部 5 は、音片データが表す音声の鼻濁音化や無声化の有無を、話速変換部 9 より供給されたピッチ成分データに基づいて判別すればよい。

なお、音片編集部 5 は、自ら設定した条件に合致する音片データが 1 個の音片につき複数あった場合は、これら複数の音片データを、設定した条件より厳格な条件に従って 1 個に絞り込むものとする。

具体的には、例えば、設定した条件が照合レベルデータの値「1」に相当するものであって、該当する音片データが複数あった場合は、照合レベルデータの値「2」に相当する検索条件にも合致するものを選択し、なお複数の音片データが選択された場合は、選択結果のうちから照合レベルデータの値「3」に相当する検索条件にも合致するものを更に選択する、等の操作を行う。照合レベルデータの値「3」に相当する検索条件で絞り込んでなお複数の音片データが残る場合は、残ったものを任意の基準で 1 個に絞り込めばよい。

一方、音片編集部 5 は、話速変換部 9 より欠落部分識別データも供給されている場合には、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出して音響処理部 41 に供給し、この音片の波形を合成するよう指示する。

指示を受けた音響処理部 41 は、音片編集部 5 より供給された表音文字列を、配信文字列データが表す表音文字列と同様に扱う。この結果、この表音文字列に含まれる表音文字が示す音声の波形を表

す圧縮波形データが検索部 4 2 により索出され、この圧縮波形データが伸長部 4 3 により元の波形データへと復元され、検索部 4 2 を介して音響処理部 4 1 へと供給される。音響処理部 4 1 は、この波形データを音片編集部 5 へと供給する。

音片編集部 5 は、音響処理部 4 1 より波形データを返送されると、この波形データと、話速変換部 9 より供給された音片データのうち音片編集部 5 が選択したものとを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する。

なお、話速変換部 9 より供給されたデータに欠落部分識別データが含まれていない場合は、音響処理部 4 1 に波形の合成を指示することなく直ちに、音片編集部 5 が選択した音片データを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力すればよい。

以上説明した、この発明の第 1 の実施の形態の音声合成システムでは、音素より大きな単位であり得る音片の波形を表す音片データが、韻律の予測結果に基づいて、録音編集方式により自然につなが合わせられ、定型メッセージを読み上げる音声合成される。音片データベース 7 の記憶容量は、音素毎に波形を記憶する場合に比べて小さくでき、また、高速に検索できる。このため、この音声合成システムは小型軽量に構成することができ、また高速な処理にも追従できる。

なお、この音声合成システムの構成は上述のものに限られない。

例えば、波形データや音片データは P C M 形式のデータである必

要はなく、データ形式は任意である。

また、波形データベース 44 や音片データベース 7 は波形データや音片データを必ずしもデータ圧縮された状態で記憶している必要はない。波形データベース 44 や音片データベース 7 が波形データや音片データをデータ圧縮されていない状態で記憶している場合、本体ユニット M1 は伸長部 43 を備えている必要はない。

また、波形データベース 44 は、必ずしも単位音声を個々に分解された形で記憶している必要はなく、例えば、複数の単位音声からなる音声の波形と、この波形内で個々の単位音声が占める位置を識別するデータとを記憶するようにしてもよい。またこの場合、音片データベース 7 が波形データベース 44 の機能を行ってもよい。つまり、波形データベース 44 内には、音片データベース 7 と同様の形式で一連の音声データが連なって記憶されていてもよく、この場合は、波形データベースとして利用するために、音声データ内の各音素毎に、表音文字やピッチ情報等が関連づけて記憶されているものとする。

また、音片データベース作成部 11 は、図示しない記録媒体ドライブ装置にセットされた記録媒体から、この記録媒体ドライブ装置を介して、音片データベース 7 に追加する新たな圧縮音片データのマテリアルとなる音片データや表音文字列を読み取ってもよい。

また、音片登録ユニット R は、必ずしも収録音片データセット記憶部 10 を備えている必要はない。

また、ピッチ成分データは音片データが表す音片のピッチ長の時間変化を表すデータであってもよい。この場合、音片編集部 5 は、ピッチ長が最も短い位置（つまり、周波数が高くなる位置）を

ピッチ成分データに基づいて特定し、この位置をアクセントの位置であると解釈すればよい。

また、音片編集部 5 は、特定の音片の韻律を表す韻律登録データをあらかじめ記憶し、定型メッセージにこの特定の音片が含まれている場合は、この韻律登録データが表す韻律を、韻律予測の結果として扱うようにしてもよい。

また、音片編集部 5 は、過去の韻律予測の結果を韻律登録データとして新たに記憶するようにしてもよい。

また、音片データベース作成部 11 は、マイクロフォン、増幅器、サンプリング回路、A/D (Analog-to-Digital) コンバータ及び PCM エンコーダなどを備えていてもよい。この場合、音片データベース作成部 11 は、収録音片データセット記憶部 10 より音片データを取得する代わりに、自己のマイクロフォンが集音した音声を表す音声信号を増幅し、サンプリングして A/D 変換した後、サンプリングされた音声信号に PCM 変調を施すことにより、音片データを作成してもよい。

また、音片編集部 5 は、音響処理部 41 より返送された波形データを話速変換部 9 に供給することにより、当該波形データが表す波形の時間長を、発声スピードデータが示すスピードに合致させるようにしてもよい。

また、音片編集部 5 は、例えば、言語処理部 1 と共にフリーテキストデータを取得し、このフリーテキストデータが表すフリーテキストに含まれる音声（表音文字列）の少なくとも一部に合致する音片データを、定型メッセージの音片データの選択処理と実質的に同一の処理を行うことによって選択して、音声の合成に用いてもよい。

この場合、音響処理部 4 1 は、音片編集部 5 が選択した音片については、この音片の波形を表す波形データを検索部 4 2 に索出させなくてもよい。なお、音片編集部 5 は、音響処理部 4 1 が合成しなくてよい音片を音響処理部 4 1 に通知し、音響処理部 4 1 はこの通知に応答して、この音片を構成する単位音声の波形の検索を中止するようにすればよい。

また、音片編集部 5 は、例えば、音響処理部 4 1 と共に配信文字列データを取得し、この配信文字列データが表す配信文字列に含まれる表音文字列を表す音片データを、定型メッセージの音片データの選択処理と実質的に同一の処理を行うことによって選択して、音声の合成に用いてもよい。この場合、音響処理部 4 1 は、音片編集部 5 が選択した音片データが表す音片については、この音片の波形を表す波形データを検索部 4 2 に索出させなくてもよい。

(第 2 の実施の形態)

次に、この発明の第 2 の実施の形態を説明する。第 3 図は、この発明の第 2 の実施の形態に係る音声合成システムの構成を示す図である。図示するように、この音声合成システムも、第 1 の実施の形態におけるものと同様、本体ユニット M 2 と、音片登録ユニット R とにより構成されている。このうち、音片登録ユニット R の構成は、第 1 の実施の形態におけるものと実質的に同一の構成を有している。

本体ユニット M 2 は、言語処理部 1 と、一般単語辞書 2 と、ユーザ単語辞書 3 と、規則合成処理部 4 と、音片編集部 5 と、検索部 6 と、音片データベース 7 と、伸長部 8 と、話速変換部 9 とにより構成されている。このうち、言語処理部 1、一般単語辞書 2、ユーザ単語辞書 3 及び音片データベース 7 は、第 1 の実施の形態における

ものと実質的に同一の構成を有している。

言語処理部 1、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 は、いずれも、CPU や DSP 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。なお、言語処理部 1、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6 及び話速変換部 9 の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。

規則合成処理部 4 は、第 1 の実施の形態におけるものと同様、音響処理部 4 1 と、検索部 4 2 と、伸長部 4 3 と、波形データベース 4 4 とより構成されている。このうち、音響処理部 4 1、検索部 4 2 及び伸長部 4 3 はいずれも、CPU や DSP 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。

なお、音響処理部 4 1、検索部 4 2 及び伸長部 4 3 の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。また、言語処理部 1、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 の一部又は全部の機能を行うプロセッサが、更に音響処理部 4 1、検索部 4 2 及び伸長部 4 3 の一部又は全部の機能を行うようにしてもよい。従って、例えば、伸長部 8 が規則合成処理部 4 の伸長部 4 3 の機能を兼ねて行うようにしてもよい。

波形データベース 4 4 は、PROM やハードディスク装置等の不揮発性メモリより構成されている。波形データベース 4 4 には、表音文字と、この表音文字が表す音素を構成する素片（すなわち、1 個の音素を構成する音声の波形 1 サイクル分（又はその他所定数のサイクル分）の音声）を表す素片波形データをエントロピー符号化

して得られる圧縮波形データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。なお、エントロピー符号化される前の素片波形データは、例えば、PCMされたデジタル形式のデータからなっていればよい。

音片編集部 5 は、一致音片決定部 5 1 と、韻律予測部 5 2 と、出力合成部 5 3 とより構成されている。一致音片決定部 5 1、韻律予測部 5 2 及び出力合成部 5 3 はいずれも、CPU や DSP 等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。

なお、一致音片決定部 5 1、韻律予測部 5 2 及び出力合成部 5 3 の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。また、言語処理部 1、音響処理部 4 1、検索部 4 2、伸長部 4 3、検索部 4 2、伸長部 4 3、音片編集部 5、検索部 6、伸長部 8 及び話速変換部 9 の一部又は全部の機能を行うプロセッサが、更に一致音片決定部 5 1、韻律予測部 5 2 及び出力合成部 5 3 の一部又は全部の機能を行うようにしてもよい。従って、例えば、出力合成部 5 3 の機能を行うプロセッサが話速変換部 9 の機能を行うようにしてもよい。

次に、第 3 図の音声合成システムの動作を説明する。

まず、言語処理部 1 が、第 1 の実施の形態におけるものと実質的に同一のフリーテキストデータを外部から取得したとする。この場合、言語処理部 1 は、第 1 の実施の形態における処理と実質的に同一の処理を行うことにより、このフリーテキストに含まれる表意文字を表音文字へと置換する。そして、置換を行った結果得られた表音文字列を、規則合成処理部 4 の音響処理部 4 1 に供給する。

音響処理部 4 1 は、言語処理部 1 より表音文字列を供給されると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す音素を構成する素片の波形を検索するよう、検索部 4 2 に指示する。また、音響処理部 4 1 は、この表音文字列を、音片編集部 5 の韻律予測部 5 2 に供給する。

検索部 4 2 は、この指示に応答して波形データベース 4 4 を検索し、この指示の内容に合致する圧縮波形データを索出する。そして、索出された圧縮波形データを伸長部 4 3 へと供給する。

伸長部 4 3 は、検索部 4 2 より供給された圧縮波形データを、圧縮される前の素片波形データへと復元し、検索部 4 2 へと返送する。検索部 4 2 は、伸長部 4 3 より返送された素片波形データを、検索結果として音響処理部 4 1 へと供給する。

一方、音響処理部 4 1 より表音文字列を供給された韻律予測部 5 2 は、この表音文字列に、例えば第 1 の実施の形態で音片編集部 5 が行うものと同様の韻律予測の手法に基づいた解析を加えることにより、この表音文字列が表す音声の韻律の予測結果を表す韻律予測データを生成する。そして、この韻律予測データを、音響処理部 4 1 に供給する。

音響処理部 4 1 は、検索部 4 2 より素片波形データを供給され、韻律予測部 5 2 より韻律予測データを供給されると、供給された素片波形データを用いて、言語処理部 1 が供給した表音文字列に含まれるそれぞれの表音文字が表す音声の波形を表す音声波形データを生成する。

具体的には、音響処理部 4 1 は、例えば、検索部 4 2 より供給された各々の素片波形データが表す素片により構成されている音素の

時間長を、韻律予測部 5 2 より供給された韻律予測データに基づいて特定する。そして、特定した音素の時間長を、当該素片波形データが表す素片の時間長で除した値に最も近い整数を求め、当該素片波形データを、求めた整数に等しい個数分相互に結合することにより、音声波形データを生成すればよい。

なお、音響処理部 4 1 は、音声波形データが表す音声の時間長を韻律予測データに基づいて決定するのみならず、音声波形データを構成する素片波形データを加工して、音声波形データが表す音声は、当該韻律予測データが示す韻律に合致する強度やイントネーション等を有するようにしてもよい。

そして、音響処理部 4 1 は、生成された音声波形データを、言語処理部 1 より供給された表音文字列内での各表音文字の並びに従った順序で、音片編集部 5 の出力合成部 5 3 へと供給する。

出力合成部 5 3 は、音響処理部 4 1 より音声波形データを供給されると、この音声波形データを、音響処理部 4 1 より供給された順序で互いに結合し、合成音声データとして出力する。フリーテキストデータに基づいて合成されたこの合成音声は、規則合成方式の手法により合成された音声に相当する。

なお、第 1 の実施の形態の音片編集部 5 と同様、出力合成部 5 3 が合成音声データを出力する手法も任意である。従って、例えば、図示しない D/A 変換器やスピーカを介して、この合成音声データが表す合成音声を再生するようにしてもよい。また、図示しないインターフェース回路を介して外部の装置やネットワークに送出してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体へ、この記録媒体ドライブ装置を介して書き込んでもよい。また、

出力合成部 5 3 の機能を行っているプロセッサが、自ら実行している他の処理へと、合成音声データを引き渡すようにしてもよい。

次に、音響処理部 4 1 が、第 1 の実施の形態におけるものと実質的に同一の配信文字列データを取得したとする。(なお、音響処理部 4 1 が配信文字列データを取得する手法も任意であり、例えば、言語処理部 1 がフリーテキストデータを取得する手法と同様の手法で配信文字列データを取得すればよい。)

この場合、音響処理部 4 1 は、配信文字列データが表す表音文字列を、言語処理部 1 より供給された表音文字列と同様に扱う。この結果、配信文字列データが表す表音文字列に含まれる表音文字が表す音素を構成する素片を表す圧縮波形データが検索部 4 2 により索出され、圧縮される前の素片波形データが伸長部 4 3 により復元される。一方で、韻律予測部 5 2 により、配信文字列データが表す表音文字列に韻律予測の手法に基づいた解析が加えられ、この結果、この表音文字列が表す音声の韻律の予測結果を表す韻律予測データが生成される。そして音響処理部 4 1 が、配信文字列データが表す表音文字列に含まれるそれぞれの表音文字が表す音声の波形を表す音声波形データを、復元された各素片波形データと、韻律予測データとに基づいて生成し、出力合成部 5 3 は、生成された音声波形データを、配信文字列データが表す表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する。配信文字列データに基づいて合成されたこの合成音声データも、規則合成方式の手法により合成された音声を表す。

次に、音片編集部 5 の一致音片決定部 5 1 が、第 1 の実施の形態におけるものと実質的に同一の定型メッセージデータ、発声スピー

ドデータ、及び照合レベルデータを取得したとする。(なお、一致音片決定部 5 1 が定型メッセージデータや発声スピードデータや照合レベルデータを取得する手法は任意であり、例えば、言語処理部 1 がフリーテキストデータを取得する手法と同様の手法で定型メッセージデータや発声スピードデータや照合レベルデータを取得すればよい。)

定型メッセージデータ、発声スピードデータ、及び照合レベルデータが一致音片決定部 5 1 に供給されると、一致音片決定部 5 1 は、定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出するよう、検索部 6 に指示する。

検索部 6 は、一致音片決定部 5 1 の指示に応答して、第 1 の実施の形態の検索部 6 と同様に音片データベース 7 を検索し、該当する圧縮音片データと、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データとをすべて索出し、索出された圧縮波形データを伸長部 4 3 へと供給する。一方、圧縮音片データを索出できなかった音片があった場合は、該当する音片を識別する欠落部分識別データを生成する。

伸長部 4 3 は、検索部 6 より供給された圧縮音片データを、圧縮される前の音片データへと復元し、検索部 6 へと返送する。検索部 6 は、伸長部 4 3 より返送された音片データと、索出された音片読みデータ、スピード初期値データ及びピッチ成分データとを、検索結果として話速変換部 9 へと供給する。また、欠落部分識別データを生成した場合は、この欠落部分識別データも話速変換部 9 へと供給する。

一方、一致音片決定部 5 1 は、話速変換部 9 に対し、話速変換部 9 に供給された音片データを変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致するようにすることを指示する。

話速変換部 9 は、一致音片決定部 5 1 の指示に応答し、検索部 6 より供給された音片データを指示に合致するように変換して、一致音片決定部 5 1 に供給する。具体的には、例えば、検索部 6 より供給された音片データを個々の音素を表す区間へと区切り、得られたそれぞれの区間について、当該区間から、当該区間が表す音素を構成する素片を表す部分を特定して、特定された部分を（1 個もしくは複数個）複製して当該区間内に挿入したり、又は、当該区間から当該部分を（1 個もしくは複数個）除去することによって、当該区間の長さを調整することにより、この音片データ全体のサンプル数を、一致音片決定部 5 1 の指示したスピードに合致する時間長にすればよい。なお、話速変換部 9 は、各区間について、素片を表す部分を挿入又は除去する個数を、各区間が表す音素相互間の時間長の比率が実質的に変化しないように決定すればよい。こうすることにより、音素同士を単に結合して合成する場合に比べて、音声のより細かい調整が可能になる。

また、話速変換部 9 は、検索部 6 より供給された音片読みデータ及びピッチ成分データも一致音片決定部 5 1 に供給し、欠落部分識別データを検索部 6 より供給された場合は、更にこの欠落部分識別データも一致音片決定部 5 1 に供給する。

なお、発声スピードデータが一致音片決定部 5 1 に供給されていない場合、一致音片決定部 5 1 は、話速変換部 9 に対し、話速変換

部 9 に供給された音片データを変換せずに一致音片決定部 5 1 に供給するよう指示すればよく、話速変換部 9 は、この指示に応答し、検索部 6 より供給された音片データをそのまま一致音片決定部 5 1 に供給すればよい。また、話速変換部 9 に供給された音片データのサンプル数が、一致音片決定部 5 1 の指示したスピードに合致する時間長に既に合致している場合も、話速変換部 9 は、この音片データを変換せずそのまま一致音片決定部 5 1 に供給すればよい。

一致音片決定部 5 1 は、話速変換部 9 より音片データ、音片読みデータ及びピッチ成分データを供給されると、第 1 の実施の形態の音片編集部 5 と同様、照合レベルデータの値に相当する条件に従って、自己に供給された音片データのうちから、定型メッセージを構成する音片の波形に近似できる波形を表す音片データを、音片 1 個につき 1 個ずつ選択する。

ただし、一致音片決定部 5 1 は、話速変換部 9 より供給された音片データのうちから、照合レベルデータの値に相当する条件を満たす音片データを選択できない音片があった場合、該当する音片を、検索部 6 が圧縮音片データを索出できなかった音片（つまり、上述の欠落部分識別データが示す音片）とみなして扱うことを決定するものとする。

そして、一致音片決定部 5 1 は、照合レベルデータの値に相当する条件を満たすものとして選択した音片データを、出力合成部 5 3 へと供給する。

また、一致音片決定部 5 1 は、話速変換部 9 より欠落部分識別データも供給されている場合、又は、照合レベルデータの値に相当する条件を満たす音片データを選択できなかった音片があった場合に

は、欠落部分識別データが示す音片（照合レベルデータの値に相当する条件を満たす音片データを選択できなかった音片を含む）の読みを表す表音文字列を定型メッセージデータより抽出して音響処理部 4 1 に供給し、この音片の波形を合成するよう指示する。

指示を受けた音響処理部 4 1 は、一致音片決定部 5 1 より供給された表音文字列を、配信文字列データが表す表音文字列と同様に扱う。この結果、この表音文字列に含まれる表音文字が表す音素を構成する素片を表す圧縮波形データが検索部 4 2 により索出され、圧縮される前の素片波形データが伸長部 4 3 により復元される。一方で、韻律予測部 5 2 により、この表音文字列が表す音片の韻律の予測結果を表す韻律予測データが生成される。そして音響処理部 4 1 が、この表音文字列に含まれるそれぞれの表音文字が表す音声の波形を表す音声波形データを、復元された各素片波形データと、韻律予測データとに基づいて生成し、生成された音声波形データを、出力合成部 5 3 へと供給する。

なお、一致音片決定部 5 1 は、韻律予測部 5 2 が既に生成して一致音片決定部 5 1 に供給した韻律予測データのうち、欠落部分識別データが示す音片に相当する部分を音響処理部 4 1 に供給するようにしてもよく、この場合、音響処理部 4 1 は、改めて韻律予測部 5 2 に当該音片の韻律予測を行わせる必要はない。このようにすれば、音片等の細かい単位毎に韻律予測を行う場合に比べて、より自然な発話が可能になる。

出力合成部 5 3 は、一致音片決定部 5 1 より音片データを供給され、音響処理部 4 1 より、素片波形データより生成された音声波形データを供給されると、供給されたそれぞれの音声波形データに含

まれる素片波形データの個数を調整することにより、当該音声波形データが表す音声の時間長を、一致音片決定部 5 1 より供給された音片データが表す音片の発声スピードと整合するようにする。

具体的には、出力合成部 5 3 は、例えば、一致音片決定部 5 1 より音片データに含まれる上述の各区間が表す音素の時間長が元の時間長に対して増減した比率を特定し、音響処理部 4 1 より供給された音声波形データが表す音素の時間長が当該比率で変化するように、各音声波形データ内の素片波形データの個数を増加あるいは減少させればよい。なお、出力合成部 5 3 は、当該比率を特定するため、例えば、一致音片決定部 5 1 が供給した音片データの生成に用いられた元の音片データを検索部 6 より取得し、これら 2 個の音片データ内で互いに同一の音素を表す区間を 1 個ずつ特定すればよい。そして、一致音片決定部 5 1 が供給した音片データ内で特定した区間内に含まれる素片の個数が、検索部 6 より取得した音片データ内で特定した区間内に含まれる素片の個数に対して増減した比率を、音素の時間長の増減の比率として特定するようにすればよい。なお、音声波形データが表す音素の時間長が、一致音片決定部 5 1 より供給された音片データが表す音片のスピードに既に整合している場合、出力合成部 5 3 は、音声波形データ内の素片波形データの個数を調整する必要はない。

そして、出力合成部 5 3 は、素片波形データの個数の調整が完了した音声波形データと、一致音片決定部 5 1 より供給された音片データとを、定型メッセージデータが示す定型メッセージ内での各音片ないし音素の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する。

なお、話速変換部 9 より供給されたデータに欠落部分識別データが含まれていない場合は、音響処理部 4 1 に波形の合成を指示することなく直ちに、音片編集部 5 が選択した音片データを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力すればよい。

以上説明した、この発明の第 2 の実施の形態の音声合成システムでも、音素より大きな単位であり得る音片の波形を表す音片データが、韻律の予測結果に基づいて、録音編集方式により自然につなが合わせられ、定型メッセージを読み上げる音声合成される。

一方、適切な音片データを選択することができなかった音片は、音素より小さな単位である素片を表す圧縮波形データを用いて、規則合成方式の手法に従って合成される。圧縮波形データが素片の波形を表すものであるため、波形データベース 4 4 の記憶容量は、圧縮波形データが音素の波形を表すものである場合に比べて小さくでき、また、高速に検索できる。このため、この音声合成システムは小型軽量に構成することができ、また高速な処理にも追従できる。

また、素片を用いて規則合成を行えば、音素を用いて規則合成を行う場合と異なり、音素の端の部分に現れる特殊な波形の影響を受けることなく音声合成を行うことができるため、少ない種類の素片で自然な音声を得ることができる。

すなわち、人が発声する音声では、先行する音素から後続の音素へと遷移する境界で、これらの音素双方の影響を受けた特殊な波形が現れることが知られており、一方、規則合成に用いられる音素は、採取した段階で既にその端部にこの特殊な波形を含んでいるため、

音素を用いて規則合成を行う場合は、音素間の境界の波形の様々なパターンを再現可能とするために膨大な種類の音素を用意するか、あるいは、音素間の境界の波形が自然な音声とは異なった合成音声を合成することで満足する必要がある。しかし、素片を用いて規則合成を行う場合は、音素の端部以外の部分から素片を採取するようにすれば、音素間の境界の特殊な波形の影響をあらかじめ排除することができる。このため、膨大な種類の素片を用意することを要せず、自然な音声を得ることができる。

なお、この発明の第2の実施の形態の音声合成システムの構成も、上述のものに限られない。

例えば、素片波形データはPCM形式のデータである必要はなく、データ形式は任意である。また、波形データベース44は素片波形データや音片データを必ずしもデータ圧縮された状態で記憶している必要はない。波形データベース44が素片波形データをデータ圧縮されていない状態で記憶している場合、本体ユニットM2は伸長部43を備えている必要はない。

また、波形データベース44は、必ずしも素片の波形を個々に分解された形で記憶している必要はなく、例えば、複数の素片からなる音声の波形と、この波形内で個々の素片が占める位置を識別するデータとを記憶するようにしてもよい。またこの場合、音片データベース7が波形データベース44の機能を行ってもよい。

また、一致音片決定部51は、第1の実施の形態の音片編集部5と同様に韻律登録データをあらかじめ記憶し、定型メッセージにこの特定の音片が含まれている場合にこの韻律登録データが表す韻律を韻律予測の結果として扱うようにしてもよく、また、過去の韻律

予測の結果を韻律登録データとして新たに記憶するようにしてもよい。

また、一致音片決定部 5 1 は、第 1 の実施の形態の音片編集部 5 と同様にフリーテキストデータや配信文字列データを取得し、これらが表すフリーテキストや配信文字列に含まれる音片の波形に近い波形を表す音片データを、定型メッセージに含まれる音片の波形に近い波形を表す音片データを選択する処理と実質的に同一の処理を行うことによって選択して、音声の合成に用いてもよい。この場合、音響処理部 4 1 は、一致音片決定部 5 1 が選択した音片データが表す音片については、この音片の波形を表す波形データを検索部 4 2 に索出させなくてもよく、また、一致音片決定部 5 1 は、音響処理部 4 1 が合成しなくてよい音片を音響処理部 4 1 に通知し、音響処理部 4 1 はこの通知に応答して、この音片を構成する単位音声の波形の検索を中止するようにすればよい。

波形データベース 4 4 が記憶する圧縮波形データは、必ずしも素片を表すものである必要はなく、例えば、第 1 の実施の形態と同様、波形データベース 4 4 が記憶する表音文字が表す単位音声の波形を表す波形データ、あるいは当該波形データをエントロピー符号化して得られるデータであってもよい。

また、波形データベース 4 4 は、素片の波形を表すデータと、音素の波形を表すデータとを、両方記憶していてもよい。この場合、音響処理部 4 1 は、配信文字列等に含まれる表音文字が表す音素のデータを検索部 4 2 に索出させ、該当する音素が索出されなかった表音文字について、当該表音文字が表す音素を構成する素片を表すデータを検索部 4 2 に索出させ、索出された、素片を表すデータを

用いて、音素を表すデータを生成するようにしてもよい。

また、話速変換部 9 が、音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる手法は任意である。従って、話速変換部 9 は、例えば第 1 の実施の形態における処理と同様に、検索部 6 より供給された音片データをリサンプリングして、この音片データのサンプル数を、一致音片決定部 5 1 の指示した発声スピードに合致する時間長に相当する数へと増減させてもよい。

また、本体ユニット M 2 は必ずしも話速変換部 9 を備えている必要はない。本体ユニット M 2 が話速変換部 9 を備えない場合、韻律予測部 5 2 が発話スピードを予測し、一致音片決定部 5 1 は、検索部 6 が取得した音片データのうち、所定の判別条件下で発話スピードが韻律予測部 5 2 による予測の結果に合致するものを選択し、一方、発話スピードが当該予測の結果に合致しないものを選択の対象から除外するものとしてもよい。なお、音片データベース 7 は、音片の読みが共通で発話スピードが互いに異なる複数の音片データを記憶していてもよい。

また、出力合成部 5 3 が、音声波形データが表す音素の時間長を、音片データが表す音片の発声スピードと整合させる手法も任意である。従って、出力合成部 5 3 は、例えば、一致音片決定部 5 1 より音片データに含まれる各区間が表す音素の時間長が元の時間長に対して増減した比率を特定した上、音声波形データをリサンプリングして、音声波形データのサンプル数を、一致音片決定部 5 1 の指示した発声スピードと整合する時間長に相当する数へと増減させてもよい。

また、発声スピードは音片毎に異なってもよい。(従って、発

声スピードデータは、音片毎に異なる発声スピードを指定するものであってもよい。)そして、出力合成部 5 3 は、互いに発声スピードが異なる 2 個の音片の間に位置する各音声の音声波形データについては、当該 2 個の音片の発声スピードを補間（例えば、直線補間）することにより、当該 2 個の音片の間にあるこれらの音声の発声スピードを決定し、決定した発声スピードに合致するように、これらの音声を表す音声波形データを変換するようにしてもよい。

また、出力合成部 5 3 は、音響処理部 4 1 より返送された音声波形データが、フリーテキストや配信文字列を読み上げる音声を構成する音声を表すものであっても、これらの音声波形データを変換して、これらの音声の時間長を、例えば一致音片決定部 5 1 に供給されている発声スピードデータが示すスピードに合致させるようにしてもよい。

また、上述のシステムでは、例えば韻律予測部 5 2 が、文章全体に対して韻律予測(発話スピードの予測も含む)を行ってもよいし、所定の単位ごとに韻律予測を行ってもよい。また、文章全体に対して韻律予測を行った場合、読みが一致する音片があれば更に韻律が所定条件内で一致するか否かを判別し、一致していれば当該音片を採用するようにしてもよい。一致する音片が存在しなかった部分については、規則合成処理部 4 が素片を基に音声を生成するものとし、ただし、素片を基に合成する部分のピッチやスピードを、文章全体若しくは所定の単位ごとに行われた韻律予測の結果に基づいて調整するものとしてもよい。これによって、音片と、素片を基に生成する音声とを組み合わせる場合でも、自然な発話が行われる。

また、言語処理部 1 に入力される文字列が表音文字列である場合、

言語処理部 1 は、韻律予測とは別に公知の自然言語解析処理を行い、一致音片決定部 5 1 が、自然言語解析処理の結果に基づいて音片の選択を行ってもよい。これによって、単語（名詞や動詞等の品詞）毎に文字列を解釈した結果を用いて音片選択を行うことが可能になり、単に表音文字列と一致する音片を選択する場合に比べて自然な発話を行うことができる。

以上、この発明の実施の形態を説明したが、この発明にかかる音声合成装置は、専用のシステムによらず、通常のコンピュータシステムを用いて実現可能である。

例えば、パーソナルコンピュータに上述の言語処理部 1、一般単語辞書 2、ユーザ単語辞書 3、音響処理部 4 1、検索部 4 2、伸長部 4 3、波形データベース 4 4、音片編集部 5、検索部 6、音片データベース 7、伸長部 8 及び話速変換部 9 の動作を実行させるためのプログラムを格納した記録媒体（CD-ROM、MO、フロッピー（登録商標）ディスク等）から該プログラムをインストールすることにより、上述の処理を実行する本体ユニット M 1 を構成することができる。

また、パーソナルコンピュータに上述の収録音片データセット記憶部 1 0、音片データベース作成部 1 1 及び圧縮部 1 2 の動作を実行させるためのプログラムを格納した媒体から該プログラムをインストールすることにより、上述の処理を実行する音片登録ユニット R を構成することができる。

そして、これらのプログラムを実行し本体ユニット M 1 や音片登録ユニット R として機能するパーソナルコンピュータが、第 1 図の音声合成システムの動作に相当する処理として、第 4 図～第 6 図に

示す処理を行うものとする。

第 4 図は、このパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

第 5 図は、このパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

第 6 図は、このパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

すなわち、このパーソナルコンピュータが、外部より、上述のフリーテキストデータを取得すると（第 4 図、ステップ S 1 0 1）、このフリーテキストデータが表すフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書 2 やユーザ単語辞書 3 を検索することにより特定し、この表意文字を、特定した表音文字へと置換する（ステップ S 1 0 2）。なお、このパーソナルコンピュータがフリーテキストデータを取得する手法は任意である。

そして、このパーソナルコンピュータは、フリーテキスト内の表意文字をすべて表音文字へと置換した結果を表す表音文字列が得られると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース 4 4 より検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する（ステップ S 1 0 3）。

次に、このパーソナルコンピュータは、索出された圧縮波形データを、圧縮される前の波形データへと復元し（ステップ S 1 0 4）、復元された波形データを、表音文字列内での各表音文字の並びに従

った順序で互いに結合し、合成音声データとして出力する（ステップS 1 0 5）。なお、このパーソナルコンピュータが合成音声データを出力する手法は任意である。

また、このパーソナルコンピュータが、外部より、上述の配信文字列データを任意の手法で取得すると（第5図、ステップS 2 0 1）、この配信文字列データが表す表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース4 4より検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する（ステップS 2 0 2）。

次に、このパーソナルコンピュータは、索出された圧縮波形データを、圧縮される前の波形データへと復元し（ステップS 2 0 3）、復元された波形データを、表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとしてステップS 1 0 5の処理と同様の処理により出力する（ステップS 2 0 4）。

一方、このパーソナルコンピュータが、外部より、上述の定型メッセージデータ及び発声スピードデータを任意の手法により取得すると（第6図、ステップS 3 0 1）、まず、この定型メッセージデータが表す定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出する（ステップS 3 0 2）。

また、ステップS 3 0 2では、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データも索出する。なお、1個の音片につき複数の圧縮音片データが該当する場合は、該当する圧縮音片データすべてを索出

する。一方、圧縮音片データを索出できなかった音片があった場合は、上述の欠落部分識別データを生成する。

次に、このパーソナルコンピュータは、索出された圧縮音片データを、圧縮される前の音片データへと復元する(ステップS303)。そして、復元された音片データを、上述の音片編集部5が行う処理と同様の処理により変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる(ステップS304)。なお、発声スピードデータが供給されていない場合は、復元された音片データを変換しなくてもよい。

次に、このパーソナルコンピュータは、定型メッセージデータが表す定型メッセージに韻律予測の手法に基づいた解析を加えることにより、この定型メッセージの韻律を予測する(ステップS305)。そして、音片の時間長が変換された音片データのうちから、定型メッセージを構成する音片の波形に最も近い波形を表す音片データを、上述の音片編集部5が行う処理と同様の処理を行うことにより、外部より取得した照合レベルデータが示す基準に従って、音片1個につき1個ずつ選択する(ステップS306)。

具体的には、ステップS306でこのパーソナルコンピュータは、例えば、上述した(1)～(3)の条件に従って音片データを特定する。すなわち、照合レベルデータの値が「1」である場合は、定型メッセージ内の音片と読みが合致する音片データをすべて、定型メッセージ内の音片の波形を表しているとみなす。また、照合レベルデータの値が「2」である場合は、読みを表す表音文字が合致し、更に、音片データのピッチ成分の周波数の時間変化を表すピッチ成分データの内容が定型メッセージに含まれる音片のアクセントの予

測結果に合致する場合に限り、この音片データが定型メッセージ内の音片の波形を表しているとみなす。また、照合レベルデータの値が「3」である場合は、読みを表す表音文字及びアクセントが合致し、更に、音片データが表す音声の鼻濁音化や無声化の有無が、定型メッセージの韻律の予測結果に合致している場合に限り、この音片データが定型メッセージ内の音片の波形を表しているとみなす。

なお、照合レベルデータが示す基準に合致する音片データが1個の音片につき複数あった場合は、これら複数の音片データを、設定した条件より厳格な条件に従って1個に絞り込むものとする。

一方、このパーソナルコンピュータは、欠落部分識別データを生成した場合、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出し、この表音文字列につき、音素毎に、配信文字列データが表す表音文字列と同様に扱って上述のステップS202～S203の処理を行うことにより、この表音文字列内の各表音文字が示す音声の波形を表す波形データを復元する（ステップS307）。

そして、このパーソナルコンピュータは、復元した波形データと、ステップS306で選択した音片データとを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する（ステップS308）。

また、例えば、パーソナルコンピュータに第3図の言語処理部1、一般単語辞書2、ユーザ単語辞書3、音響処理部41、検索部42、伸長部43、波形データベース44、音片編集部5、検索部6、音片データベース7、伸長部8及び話速変換部9の動作を実行させる

ためのプログラムを格納した記録媒体から該プログラムをインストールすることにより、上述の処理を実行する本体ユニットM2を構成することもできる。

そして、このプログラムを実行し本体ユニットM2として機能するパーソナルコンピュータが、第3図の音声合成システムの動作に相当する処理として、第7図～第9図に示す処理を行うようにすることもできる。

第7図は、本体ユニットM2の機能を行うパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

第8図は、本体ユニットM2の機能を行うパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

第9図は、本体ユニットM2の機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

すなわち、このパーソナルコンピュータが、外部より、上述のフリーテキストデータを取得すると（第7図、ステップS401）、このフリーテキストデータが表すフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書2やユーザ単語辞書3を検索することにより特定し、この表意文字を、特定した表音文字へと置換する（ステップS402）。なお、このパーソナルコンピュータがフリーテキストデータを取得する手法は任意である。

そして、このパーソナルコンピュータは、フリーテキスト内の表

意文字をすべて表音文字へと置換した結果を表す表音文字列が得られると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース４４より検索し、表音文字列に含まれるそれぞれの表音文字が表す音素を構成する素片の波形を表す圧縮波形データを索出し（ステップＳ４０３）、索出された圧縮波形データを、圧縮される前の素片波形データへと復元する（ステップＳ４０４）。

一方で、このパーソナルコンピュータは、フリーテキストデータに韻律予測の手法に基づいた解析を加えることにより、フリーテキストが表す音声の韻律を予測する（ステップＳ４０５）。そして、ステップＳ４０４で復元された素片波形データと、ステップＳ４０５における韻律の予測結果とに基づいて音声波形データを生成し（ステップＳ４０６）、得られた音声波形データを、表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する（ステップＳ４０７）。なお、このパーソナルコンピュータが合成音声データを出力する手法は任意である。

また、このパーソナルコンピュータが、外部より、上述の配信文字列データを任意の手法で取得すると（第８図、ステップＳ５０１）、この配信文字列データが表す表音文字列に含まれるそれぞれの表音文字について、上述のステップＳ４０３～４０４と同様に、当該表音文字が表す音素を構成する素片の波形を表す圧縮波形データを索出する処理、及び、索出された圧縮波形データを素片波形データへと復元する処理を行う（ステップＳ５０２）。

一方でこのパーソナルコンピュータは、配信文字列に韻律予測の手法に基づいた解析を加えることにより、配信文字列が表す音声の

韻律を予測し（ステップ S 5 0 3）、ステップ S 5 0 2 で復元された素片波形データと、ステップ S 5 0 3 における韻律の予測結果とに基づいて音声波形データを生成し（ステップ S 5 0 4）、得られた音声波形データを、表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとしてステップ S 4 0 7 の処理と同様の処理により出力する（ステップ S 5 0 5）。

一方、このパーソナルコンピュータが、外部より、上述の定型メッセージデータ及び発声スピードデータを任意の手法により取得すると（第 9 図、ステップ S 6 0 1）、まず、この定型メッセージデータが表す定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出する（ステップ S 6 0 2）。

また、ステップ S 6 0 2 では、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データも索出する。なお、1 個の音片につき複数の圧縮音片データが該当する場合は、該当する圧縮音片データすべてを索出する。一方、圧縮音片データを索出できなかった音片があった場合は、上述の欠落部分識別データを生成する。

次に、このパーソナルコンピュータは、索出された圧縮音片データを、圧縮される前の素片音片データへと復元する（ステップ S 6 0 3）。そして、復元された音片データを、上述の出力合成部 5 3 が行う処理と同様の処理により変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる（ステップ S 6 0 4）。なお、発声スピードデータが供給されていない場合は、復元された音片データを変換しなくてもよい。

次に、このパーソナルコンピュータは、定型メッセージデータが表す定型メッセージに韻律予測の手法に基づいた解析を加えることにより、この定型メッセージの韻律を予測する（ステップ S 6 0 5）。そして、音片の時間長が変換された音片データのうちから、定型メッセージを構成する音片の波形に最も近い波形を表す音片データを、上述の一致音片決定部 5 1 が行う処理と同様の処理を行うことにより、外部より取得した照合レベルデータが示す基準に従って、音片 1 個につき 1 個ずつ選択する（ステップ S 6 0 6）。

具体的には、ステップ S 6 0 6 でこのパーソナルコンピュータは、例えば、上述のステップ 3 0 6 の処理と同様の処理を行うことにより、上述した（1）～（3）の条件に従って音片データを特定する。なお、照合レベルデータが示す基準に合致する音片データが 1 個の音片につき複数あった場合は、これら複数の音片データを、設定した条件より厳格な条件に従って 1 個に絞り込むものとする。また、照合レベルデータの値に相当する条件を満たす音片データを選択できない音片があった場合は、該当する音片を、圧縮音片データを索出できなかった音片として扱うことと決定し、例えば欠落部分識別データを生成するものとする。

一方、このパーソナルコンピュータは、欠落部分識別データを生成した場合、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出し、この表音文字列につき、音素毎に、配信文字列データが表す表音文字列と同様に扱って上述のステップ S 5 0 2 ～ S 5 0 4 の処理と同様の処理を行うことにより、この表音文字列内の各表音文字が示す音声の波形を表す音声波形データを生成する（ステップ S 6 0 7）。

ただし、ステップ S 6 0 7 でこのパーソナルコンピュータは、ステップ S 5 0 3 の処理に相当する処理を行う代わりに、ステップ S 6 0 5 における韻律予測の結果を用いて音声波形データを生成するようにしてもよい。

次に、このパーソナルコンピュータは、上述の出力合成部 5 3 が行う処理と同様の処理を行うことにより、ステップ S 6 0 7 で生成された音声波形データに含まれる素片波形データの個数を調整し、当該音声波形データが表す音声の時間長を、ステップ S 6 0 6 で選択された音片データが表す音片の発声スピードと整合するようにする（ステップ S 6 0 8）。

すなわち、ステップ S 6 0 8 でこのパーソナルコンピュータは、例えば、ステップ S 6 0 6 で選択された音片データに含まれる上述の各区間が表す音素の時間長が元の時間長に対して増減した比率を特定し、ステップ S 6 0 7 で生成された音声波形データが表す音声の時間長が当該比率で変化するように、各音声波形データ内の素片波形データの個数を増加あるいは減少させればよい。なお、当該比率を特定するため、例えば、ステップ S 6 0 6 で選択された音片データ（発声スピード変換後の音片データ）と、当該音片データがステップ S 6 0 4 で変換を受ける前の元の音片データとの内で互いに同一の音声を表す区間を 1 個ずつ特定し、発声スピード変換後の音片データ内で特定した区間内に含まれる素片の個数が、元の音片データ内で特定した区間内に含まれる素片の個数に対して増減した比率を、音声の時間長の増減の比率として特定するようにすればよい。なお、音声波形データが表す音声の時間長が、発声スピード変換後の音片データが表す音片のスピードに既に整合している場合、この

パーソナルコンピュータは音声波形データ内の素片波形データの個数を調整する必要はない。

そして、このパーソナルコンピュータは、ステップ S 6 0 8 の処理を経た音声波形データと、ステップ S 6 0 6 で選択した音片データとを、定型メッセージデータが示す定型メッセージ内での表音文字列の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する（ステップ S 6 0 9）。

なお、パーソナルコンピュータに本体ユニット M 1 や本体ユニット M 2 や音片登録ユニット R の機能を行わせるプログラムは、例えば、通信回線の掲示板（B B S）にアップロードし、これを通信回線を介して配信してもよく、また、これらのプログラムを表す信号により搬送波を変調し、得られた変調波を伝送し、この変調波を受信した装置が変調波を復調してこれらのプログラムを復元するようにしてもよい。

そして、これらのプログラムを起動し、O S の制御下に、他のアプリケーションプログラムと同様に実行することにより、上述の処理を実行することができる。

なお、O S が処理の一部を分担する場合、あるいは、O S が本願発明の 1 つの構成要素の一部を構成するような場合には、記録媒体には、その部分を除いたプログラムを格納してもよい。この場合も、この発明では、その記録媒体には、コンピュータが実行する各機能又はステップを実行するためのプログラムが格納されているものとする。

請求の範囲

1. 音片を表す音片データを複数記憶する音片記憶手段と、
文章を表す文章情報を入力し、

各前記音片データのうちから、前記文章を構成する音声と読みが
共通している音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選
択できなかった音声について、当該音声の波形を表す音声データを
合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が
合成した音声データを互いに結合することにより、合成音声を表す
データを生成する合成手段と、

より構成されることを特徴とする音声合成装置。

2. 音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音声の韻律を
予測する韻律予測手段と、

各前記音片データのうちから、前記文章を構成する音声と読みが
共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致す
る音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選
択できなかった音声について、当該音片の波形を表す音声データを
合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が
合成した音声データを互いに結合することにより、合成音声を表す
データを生成する合成手段と、

より構成されることを特徴とする音声合成装置。

3. 前記選択手段は、韻律が韻律予測結果に前記所定の条件下で合致しない音片データを、選択の対象から除外する、

ことを特徴とする請求項2に記載の音声合成装置。

4. 前記欠落部分合成手段は、

音素を表し、又は、音素を構成する素片を表すデータを複数記憶する記憶手段と、

前記選択手段が音片データを選択できなかった前記音声に含まれる音素を特定し、特定した音素又は当該音素を構成する素片を表すデータを前記記憶手段より取得して互いに結合することにより、当該音声の波形を表す音声データを合成する合成手段と、を備える、

ことを特徴とする請求項2又は3に記載の音声合成装置。

5. 前記欠落部分合成手段は、前記選択手段が音片データを選択できなかった前記音声の韻律を予測する欠落部分韻律予測手段を備え、

前記合成手段は、前記選択手段が音片データを選択できなかった前記音声に含まれる音素を特定し、特定した音素又は当該音素を構成する素片を表すデータを前記記憶手段より取得し、取得したデータを、当該データが表す音素又は素片が、前記欠落部分韻律予測手段による韻律の予測結果に合致するように変換して、変換されたデータを互いに結合することにより、当該音声の波形を表す音声データを合成する、

ことを特徴とする請求項4に記載の音声合成装置。

6. 前記欠落部分合成手段は、前記韻律予測手段が予測した韻律に基づいて、前記選択手段が音片データを選択できなかった音声に

ついて、当該音片の波形を表す音声データを合成する、

ことを特徴とする請求項 2、3 又は 4 に記載の音声合成装置。

7. 前記音片記憶手段は、音片データが表す音片のピッチの時間変化を表す韻律データを、当該音片データに対応付けて記憶しており、

前記選択手段は、各前記音片データのうちから、前記文章を構成する音声と読みが共通しており、且つ、対応付けられている韻律データが表すピッチの時間変化が韻律の予測結果に最も近い音片データを選択する、

ことを特徴とする請求項 2 乃至 6 のいずれか 1 項に記載の音声合成装置。

8. 前記合成音声を発声するスピードの条件を指定する発声スピードデータを取得し、前記合成音声を表すデータを構成する音片データ及び／又は音声データを、当該発声スピードデータが指定する条件を満たすスピードで発話される音声を表すように選択又は変換する発話スピード変換手段を備える、

ことを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の音声合成装置。

9. 前記発話スピード変換手段は、前記合成音声を表すデータを構成する音片データ及び／又は音声データから素片を表す区間を除去し、又は、当該音片データ及び／又は音声データに素片を表す区間を追加することによって、当該音片データ及び／又は音声データを、前記発声スピードデータが指定する条件を満たすスピードで発話される音声を表すよう変換する、

ことを特徴とする請求項 8 に記載の音声合成装置。

10. 前記音片記憶手段は、音片データの読みを表す表音データを、当該音片データに対応付けて記憶しており、

前記選択手段は、前記文章を構成する音声の読みに合致する読みを表す表音データが対応付けられている音片データを、当該音声と読みが共通する音片データとして扱う、

ことを特徴とする請求項1乃至9のいずれか1項に記載の音声合成装置。

11. 音片を表す音片データを複数記憶し、

文章を表す文章情報を入力し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択し、

前記文章を構成する音声のうち、音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成し、

選択した音片データ及び合成した音声データを互いに結合することにより、合成音声を表すデータを生成する、

ことを特徴とする音声合成方法。

12. 音片を表す音片データを複数記憶し、

文章を表す文章情報を入力して、当該文章を構成する音声の韻律を予測し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択し、

前記文章を構成する音声のうち、音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成し、

選択した音片データ及び合成した音声データを互いに結合するこ

とにより、合成音声を表すデータを生成する、

ことを特徴とする音声合成方法。

13. コンピュータを、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通している音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

して機能させるためのプログラム。

14. コンピュータを、

音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に所定の条件下で合致する音片データを選択する選択手段と、

前記文章を構成する音声のうち、前記選択手段が音片データを選択できなかった音声について、当該音声の波形を表す音声データを合成する欠落部分合成手段と、

前記選択手段が選択した音片データ及び前記欠落部分合成手段が

合成した音声データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

して機能させるためのプログラム。

15. 音片を表す音片データを複数記憶する音片記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に最も近い音片データを選択する選択手段と、

選択された音片データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

より構成されることを特徴とする音声合成装置。

16. 前記選択手段は、韻律が韻律予測結果に所定の条件下で合致しない音片データを、選択の対象から除外する、

ことを特徴とする請求項15に記載の音声合成装置。

17. 前記合成音声を発声するスピードの条件を指定する発声スピードデータを取得し、前記合成音声を表すデータを構成する音片データ及び／又は音声データを、当該発声スピードデータが指定する条件を満たすスピードで発話される音声を表すように選択又は変換する発話スピード変換手段を備える、

ことを特徴とする請求項15又は16に記載の音声合成装置。

18. 前記発話スピード変換手段は、前記合成音声を表すデータを構成する音片データ及び／又は音声データから素片を表す区間を除去し、又は、当該音片データ及び／又は音声データに素片を表す区間を追加することによって、当該音片データ及び／又は音声データ

を、前記発声スピードデータが指定する条件を満たすスピードで発話される音声を表すよう変換する、

ことを特徴とする請求項 17 に記載の音声合成装置。

19. 前記音片記憶手段は、音片データが表す音片のピッチの時間変化を表す韻律データを、当該音片データに対応付けて記憶しており、

前記選択手段は、各前記音片データのうちから、前記文章を構成する音声と読みが共通しており、且つ、対応付けられている韻律データが表すピッチの時間変化が韻律の予測結果に最も近い音片データを選択する、

ことを特徴とする請求項 15 乃至 18 のいずれか 1 項に記載の音声合成装置。

20. 前記音片記憶手段は、音片データの読みを表す表音データを、当該音片データに対応付けて記憶しており、

前記選択手段は、前記文章を構成する音声の読みに合致する読みを表す表音データが対応付けられている音片データを、当該音声と読みが共通する音片データとして扱う、

ことを特徴とする請求項 15 乃至 19 のいずれか 1 項に記載の音声合成装置。

21. 音片を表す音片データを複数記憶し、

文章を表す文章情報を入力して、当該文章を構成する音声の韻律を予測し、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に最も近い音片データを選択し、

選択された音片データを互いに結合することにより、合成音声を表すデータを生成する、

ことを特徴とする音声合成方法。

22. コンピュータを、

音片を表す音片データを複数記憶する音片記憶手段と、

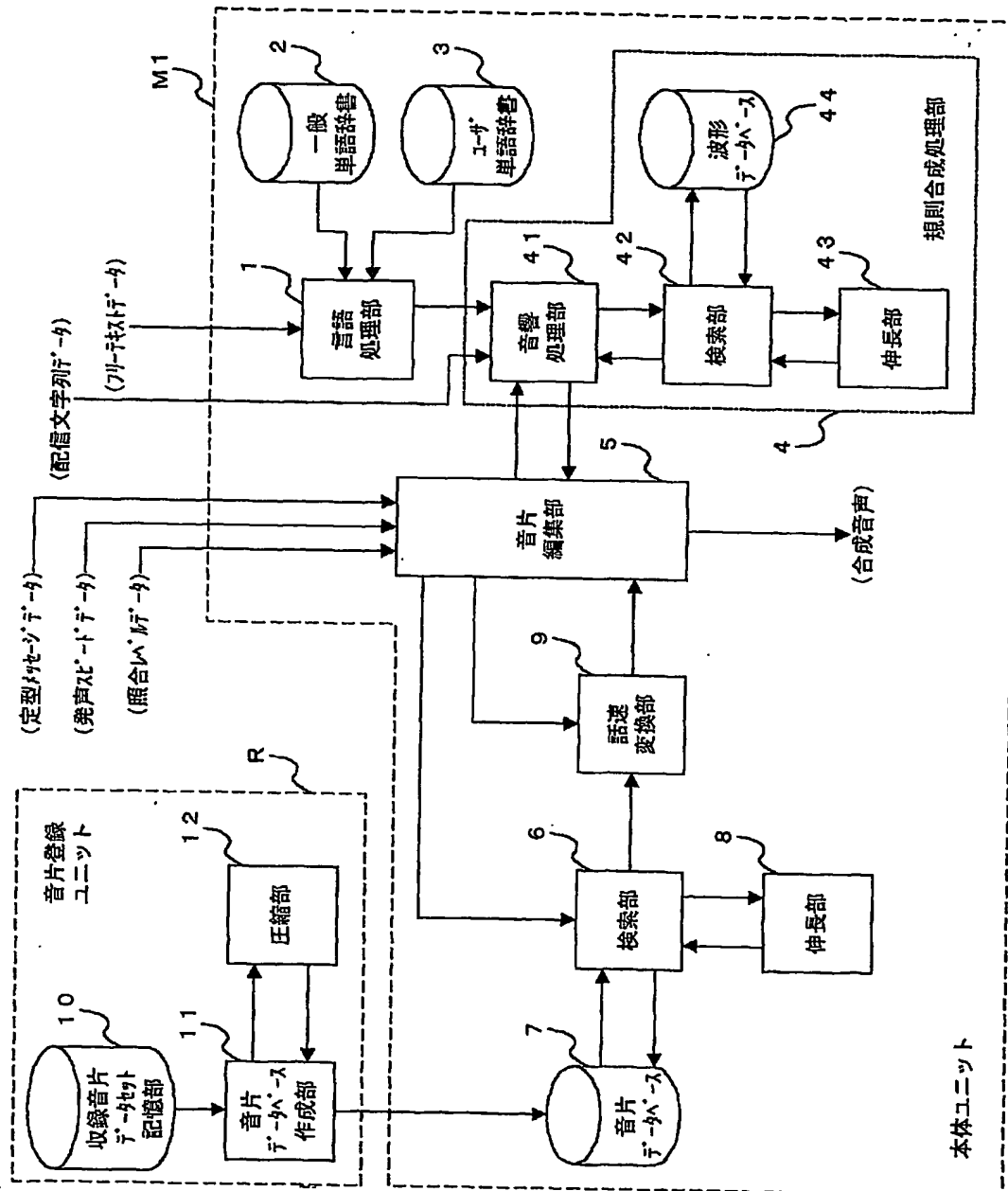
文章を表す文章情報を入力し、当該文章を構成する音声の韻律を予測する韻律予測手段と、

各前記音片データのうちから、前記文章を構成する音声と読みが共通していて、且つ、韻律が韻律予測結果に最も近い音片データを選択する選択手段と、

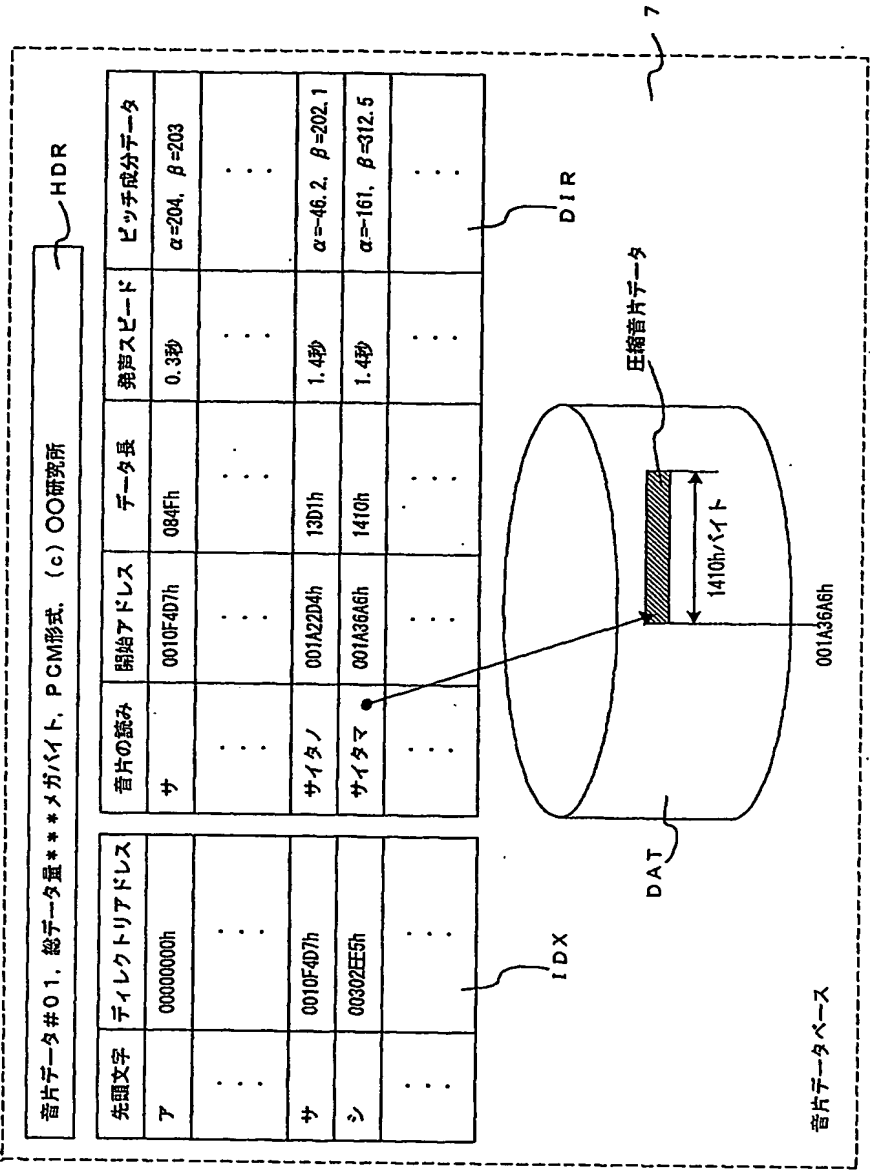
選択された音片データを互いに結合することにより、合成音声を表すデータを生成する合成手段と、

して機能させるためのプログラム。

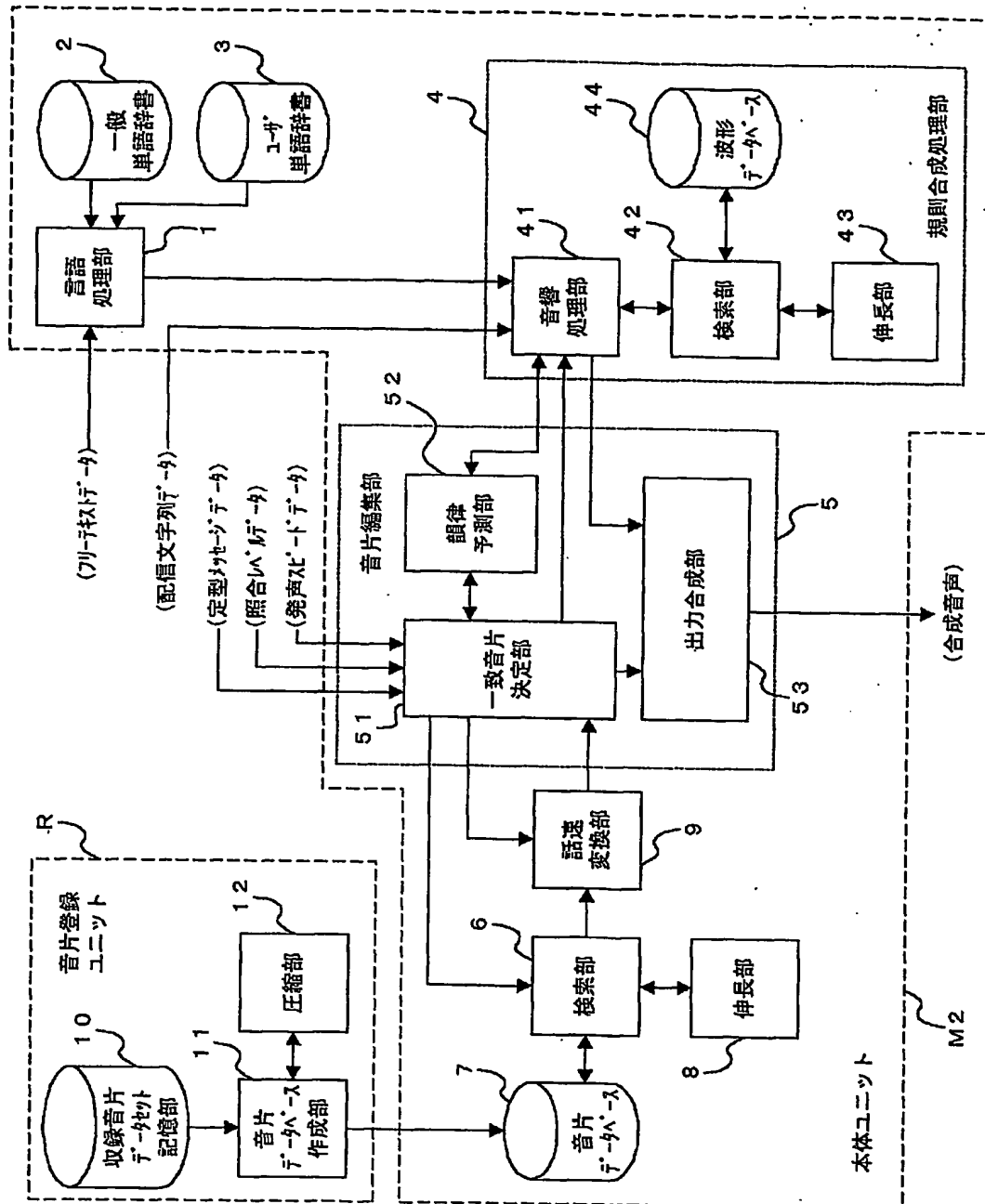
第1図



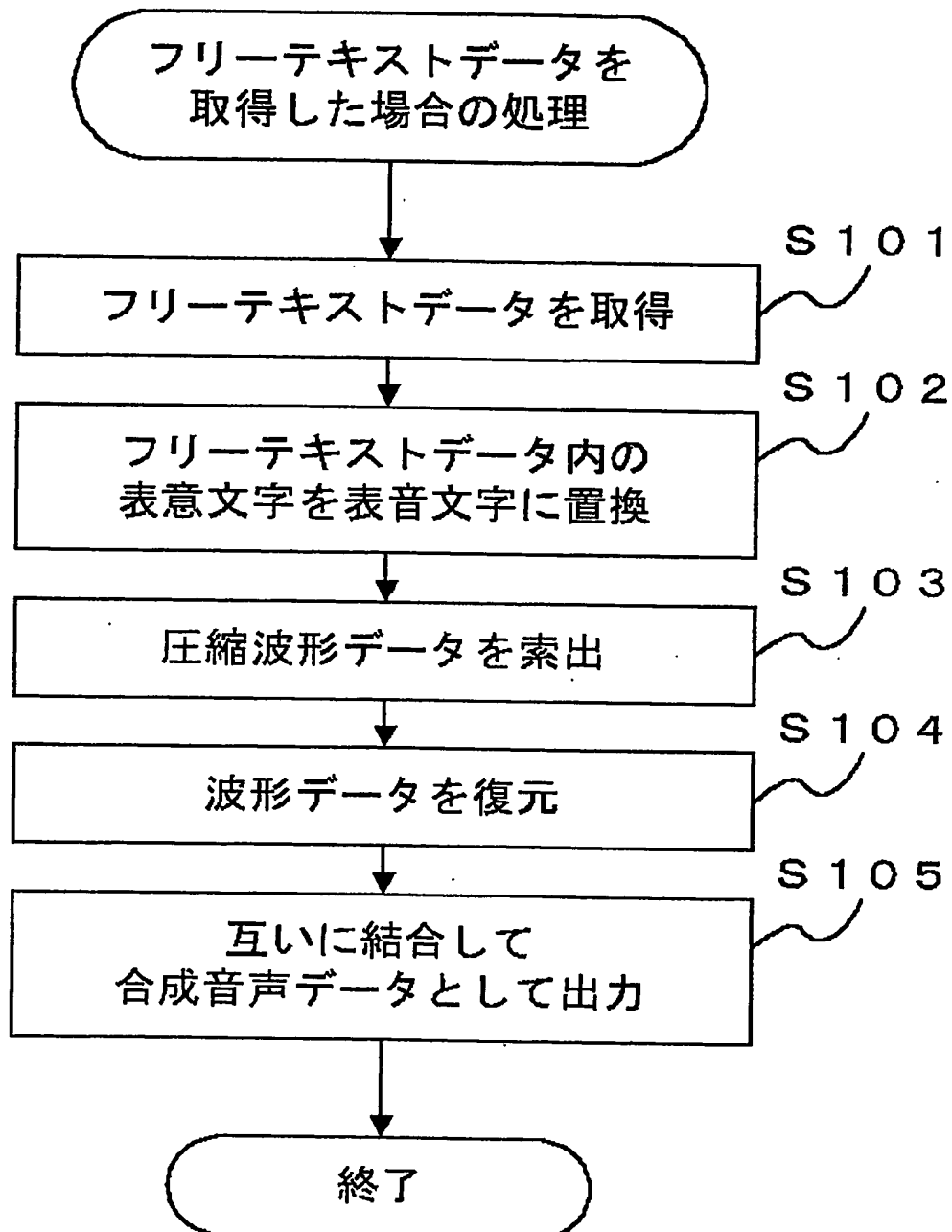
第2図



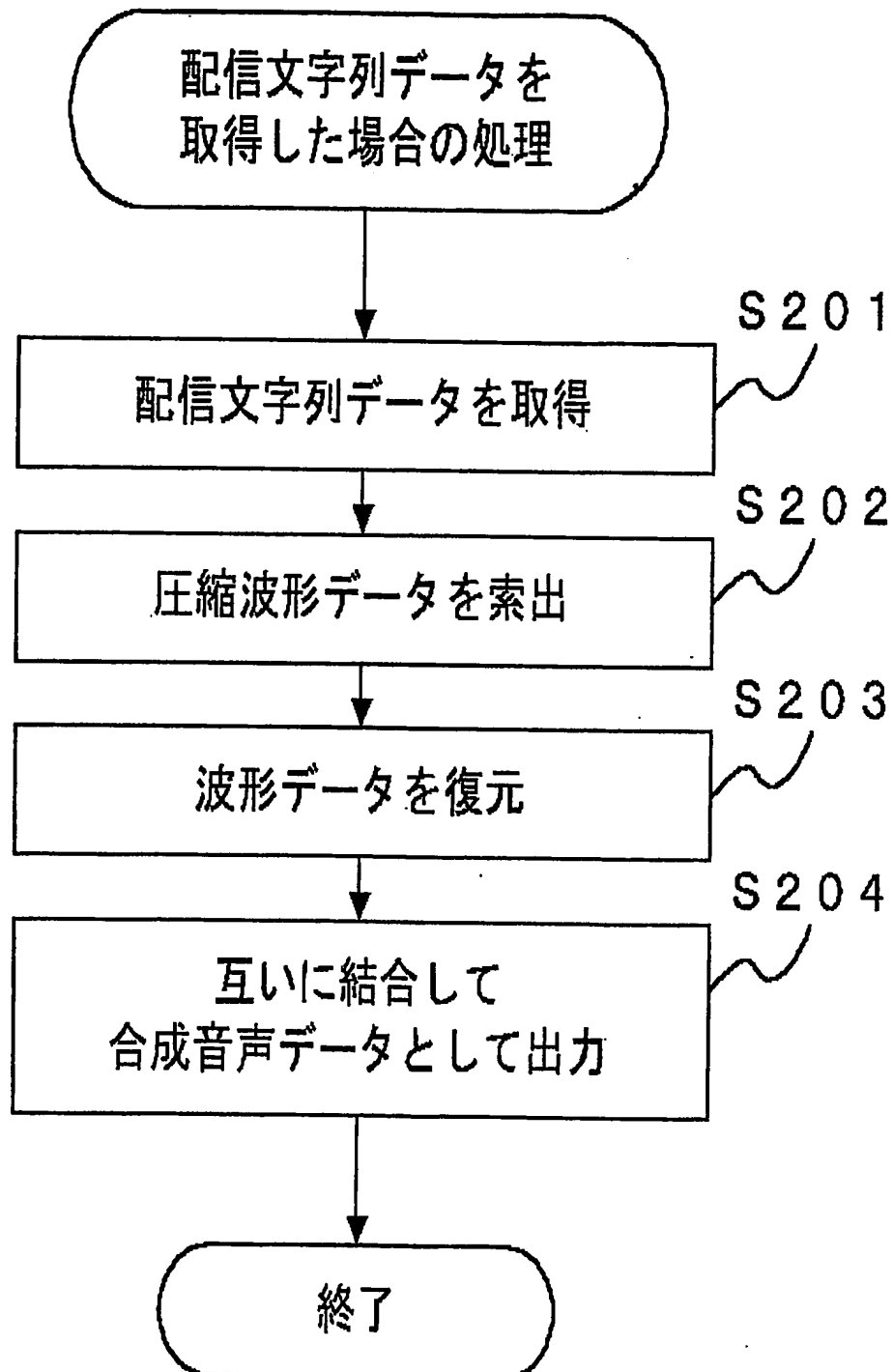
第3図



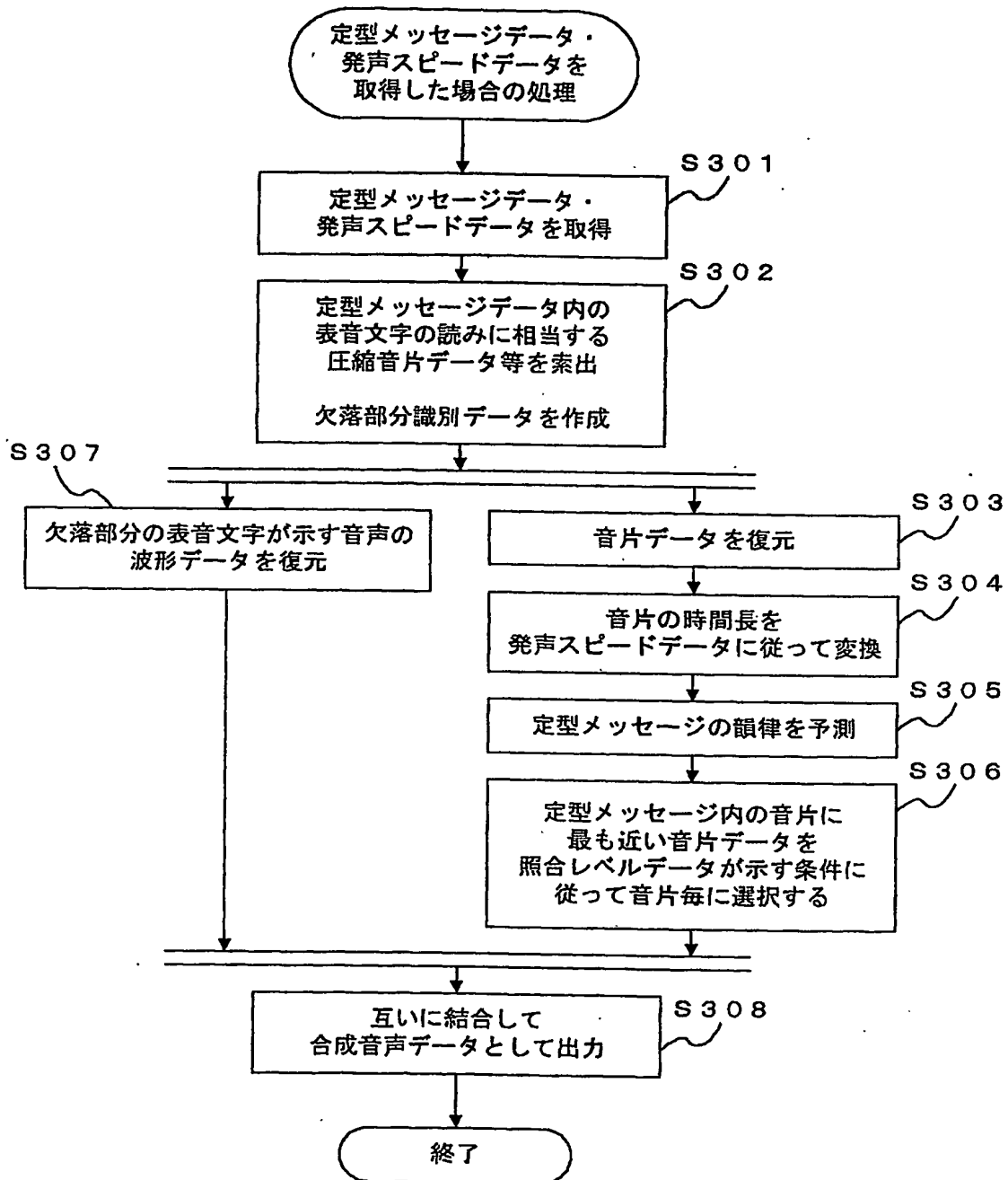
第4図



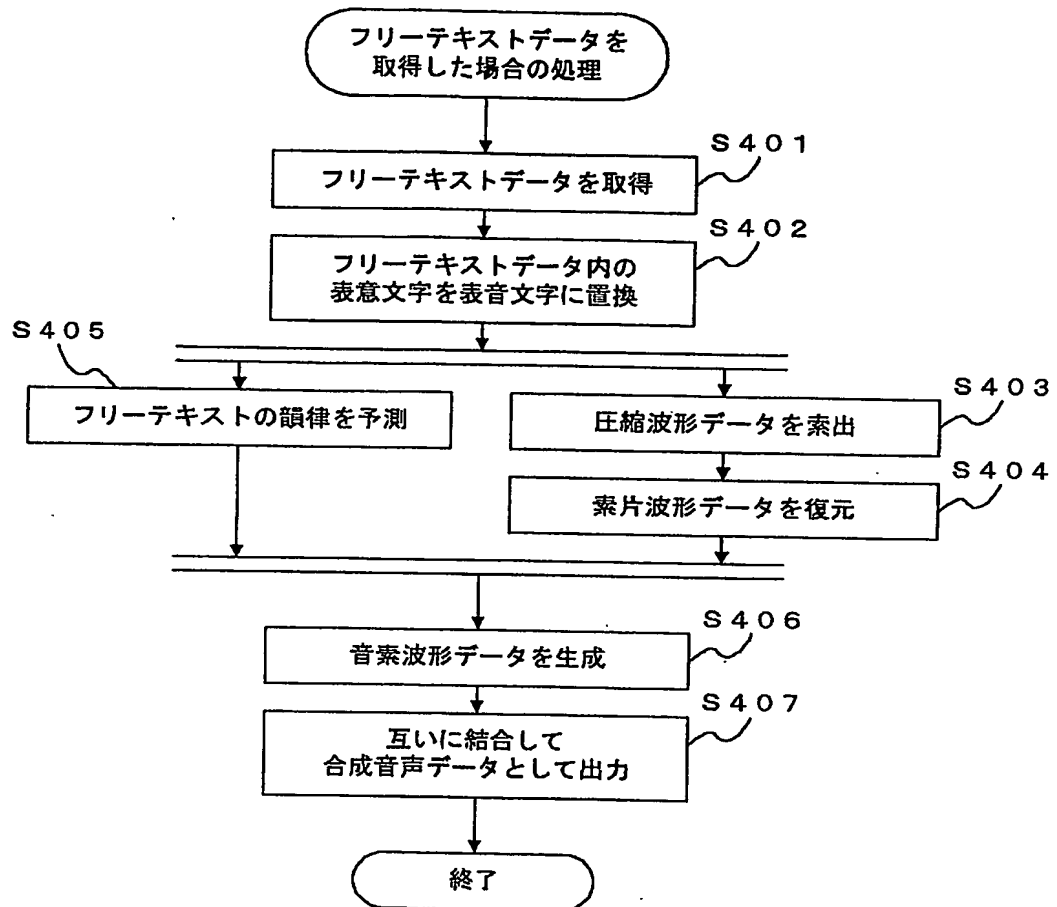
第5図



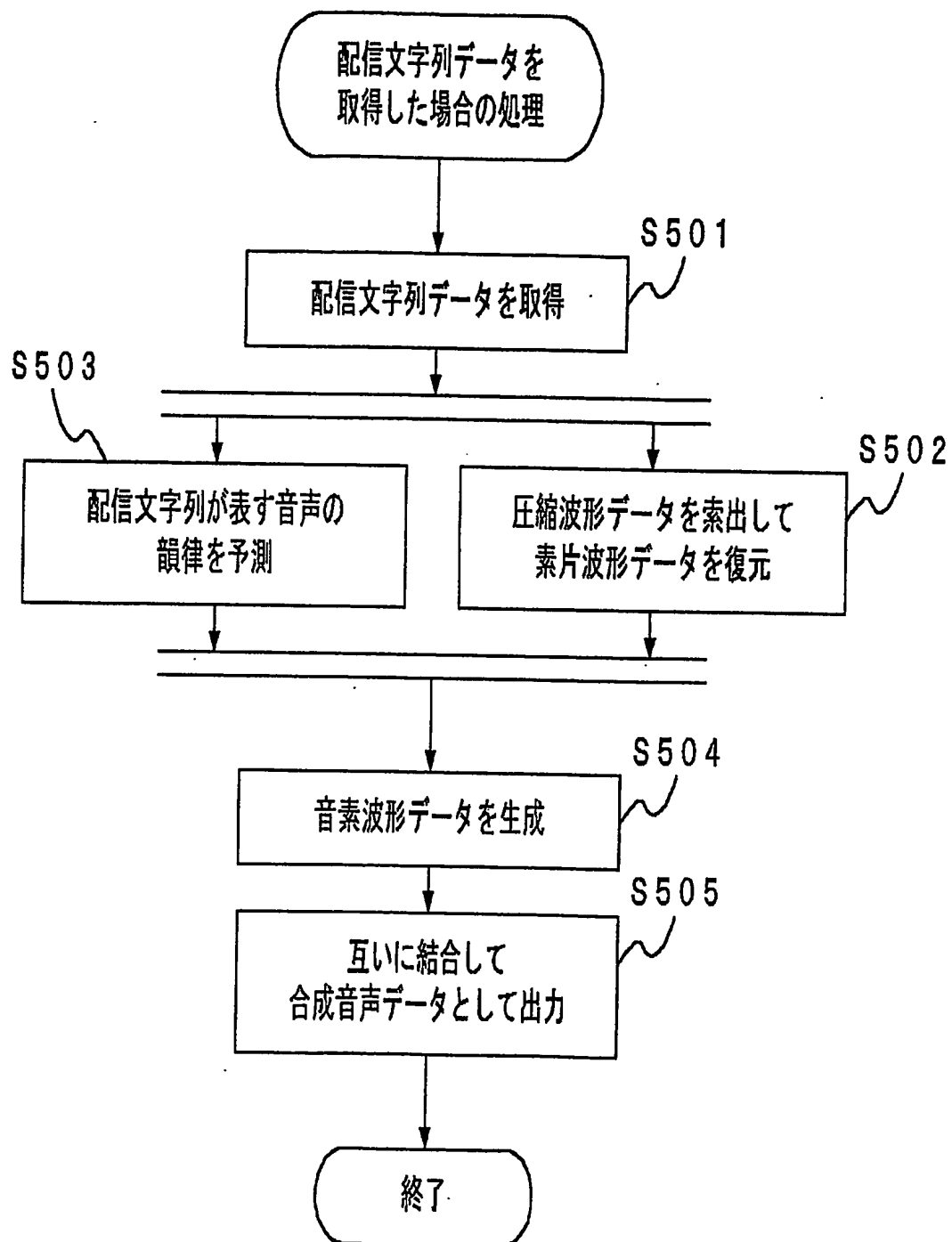
第6図



第7図

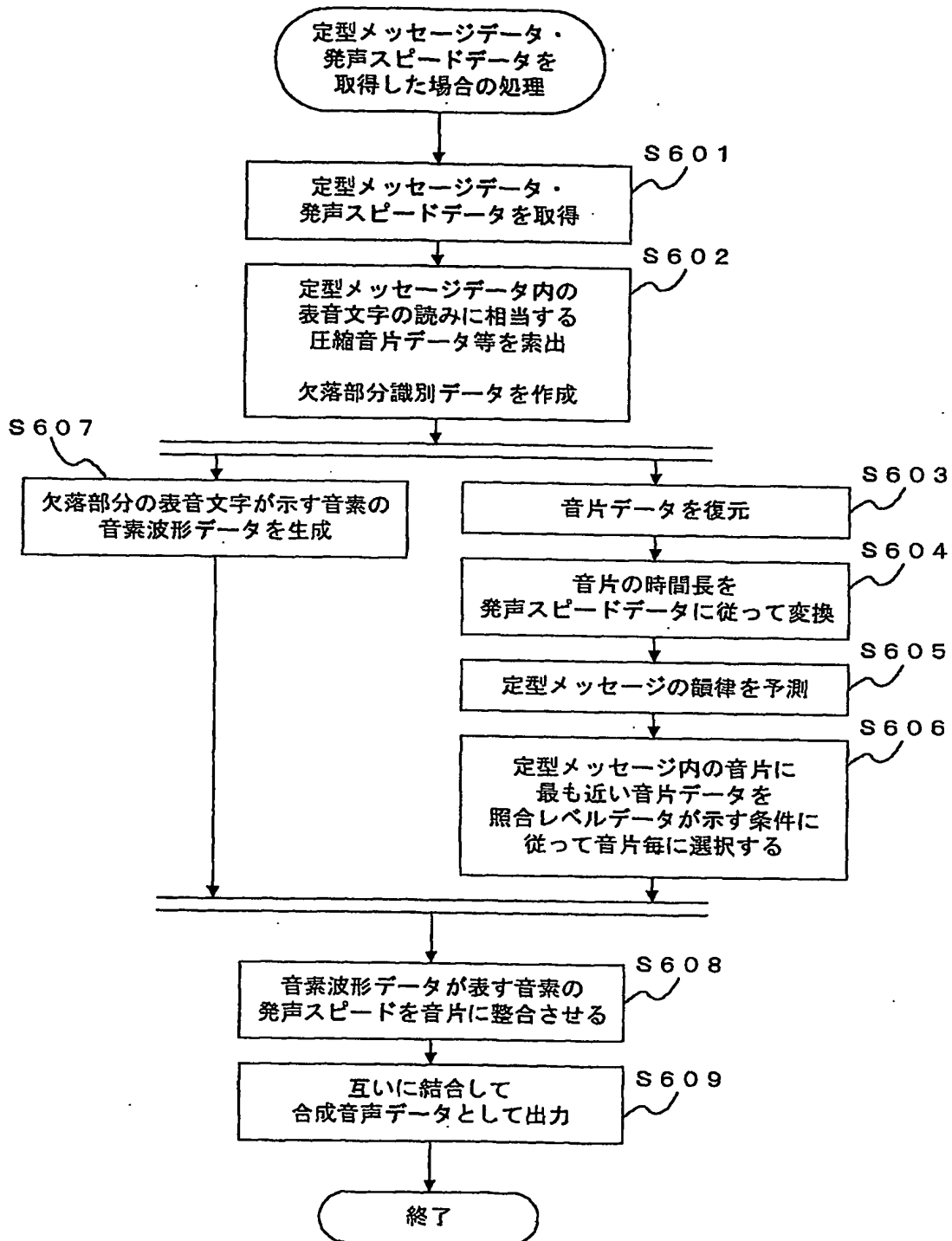


第8図



9/9

第9図



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/008087

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl⁷ G10L13/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl⁷ G10L13/00-13/08Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2004
Kokai Jitsuyo Shinan Koho 1971-2004 Toroku Jitsuyo Shinan Koho 1994-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 7-319497 A (NTT Data Communications Systems Corp.), 08 December, 1995 (08.12.95), Full text; Figs. 1 to 6 (Family: none)	1-22
A	JP 1-284898 A (Nippon Telegraph And Telephone Corp.), 16 November, 1989 (16.11.89), Full text; Figs. 1 to 3 (Family: none)	1-22
A	JP 61-59400 A (Fujitsu Ltd.), 26 March, 1986 (26.03.86), Full text; Figs. 1 to 2 (Family: none)	1-22

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search
20 August, 2004 (20.08.04)Date of mailing of the international search report
07 September, 2004 (07.09.04)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/008087

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 6-318094 A (Sharp Corp.), 15 November, 1994 (15.11.94), Full text; Figs. 1 to 2 (Family: none)	1-22
A	JP 8-87297 A (Fujitsu Ltd.), 02 April, 1996 (02.04.96), Full text; Figs. 1 to 10 (Family: none)	1-22
A	JP 9-230893 A (NTT Data Communications Systems Corp.), 05 September, 1997 (05.09.97), Full text; Figs. 1 to 9 (Family: none)	1-22
A	JP 9-319394 A (Toshiba Corp.), 12 December, 1997 (12.12.97), Full text; Figs. 1 to 13 & JP 9-319391 A & US 2003/88418 A1	1-22
A	JP 11-249676 A (Secom Co., Ltd.), 17 September, 1999 (17.09.99), Full text; Figs. 1 to 8 (Family: none)	1-22
A	JP 11-249679 A (Ricoh Co., Ltd.), 17 September, 1999 (17.09.99), Full text; Figs. 1 to 8 (Family: none)	1-22
A	JP 2003-5774 A (Matsushita Electric Industrial Co., Ltd.), 08 January, 2003 (08.01.03), Full text; Figs. 1 to 3 (Family: none)	1-22

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ G10L13/06

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ G10L13/00-13/08

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1922-1996年

日本国公開実用新案公報 1971-2004年

日本国実用新案登録公報 1996-2004年

日本国登録実用新案公報 1994-2004年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	J P 7-319497 A (エヌ・ティ・ティ・データ通信株式会社), 1995.12.08, 全文, 1-6図 (ファミリーなし)	1-22
A	J P 1-284898 A (日本電信電話株式会社) 1989.11.16, 全文, 第1-3図 (ファミリーなし)	1-22
A	J P 61-59400 A (富士通株式会社) 1986.03.26, 全文, 第1-2図 (ファミリーなし)	1-22

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの

「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」 口頭による開示、使用、展示等に言及する文献

「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」 同一パテントファミリー文献

国際調査を完了した日

20.08.2004

国際調査報告の発送日 07.9.2004

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

山下 剛史

5C

8946

電話番号 03-3581-1101 内線 3540

C (続き) 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	J P 6-318094 A (シャープ株式会社) 1994.11.15, 全文, 図1-2 (ファミリーなし)	1-22
A	J P 8-87297 A (富士通株式会社) 1996.04.02, 全文, 図1-10 (ファミリーなし)	1-22
A	J P 9-230893 A (エヌ・ティ・ティ・データ通信株式会 社), 1997.09.05, 全文, 図1-9 (ファミリーなし)	1-22
A	J P 9-319394 A (株式会社東芝) 1997.12.12, 全文, 図1-13 & JP 9-319391 A & US 2003/88418 A1	1-22
A	J P 11-249676 A (セコム株式会社) 1999.09.17, 全文, 図1-8 (ファミリーなし)	1-22
A	J P 11-249679 A (株式会社リコー) 1999.09.17, 全文, 図1-8 (ファミリーなし)	1-22
A	J P 2003-5774 A (松下電器産業株式会社) 2003.01.08, 全文, 図1-3 (ファミリーなし)	1-22